Single- and cross-disorder genome-wide studies in psychiatry and neurology

Dick Schijven

Colofon

© Dick Schijven, 2019

ISBN: 978-90-393-7097-1

Cover design: Willem Mels
Layout: Willem Mels & Dick Schijven
Print: ProefschriftMaken || www.proefschriftmaken.nl

Single- and cross-disorder genome-wide studies in psychiatry and neurology

Genoomwijde studies op het grensvlak van diagnostische classificatie in de psychiatrie en neurologie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

dinsdag 12 maart 2019 des middags te 2.30 uur

door

Dick Schijven

geboren op 14 september 1989
te Tilburg

Promotoren: Prof.dr. J.H. Veldink
Prof.dr. R.S. Kahn

Copromotor: Dr. J.J. Luykx

Table of contents

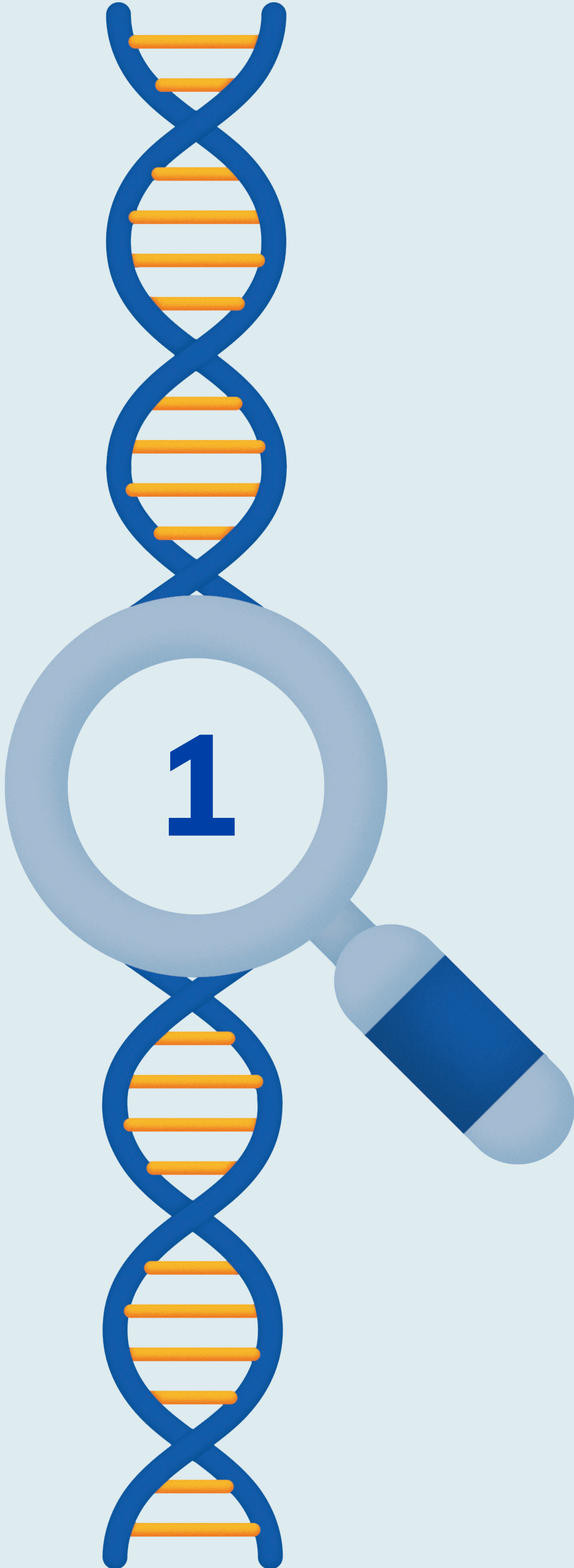
CHAPTER 1	General introduction	**7**

PART 1 | Single-disorder studies

CHAPTER 2	Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling	**27**
CHAPTER 3	Multivariate genome-wide analysis of stress-related quantitative phenotypes	**65**
CHAPTER 4	The effect of genetic vulnerability and military deployment on the development of posttraumatic stress disorder and depressive symptoms	**105**

PART 2 | Cross-disorder studies

CHAPTER 5	Genetic cross-disorder analysis in psychiatry: from methodology to clinical utility	**135**
CHAPTER 6	Genetic correlation between amyotrophic lateral sclerosis and schizophrenia	**149**
CHAPTER 7	Common genetic risk is not shared between amyotrophic lateral sclerosis and epilepsy	**183**
CHAPTER 8	Summary and general discussion	**201**
APPENDIX I	Consortium members	**231**
APPENDIX II	Samenvatting in het Nederlands	**237**
APPENDIX III	Dankwoord	**247**
APPENDIX IV	Curriculum vitae and list of publications	**255**


1

Chapter 1

General introduction

1.1 Heritability and the genetic code

The phenomenon of phenotypic variation occurring in populations and pedigrees has been the source of scientific debate for over 150 years. In the nineteenth century, the foundation for a theory on inheritance of traits was laid out by Gregor Mendel, who devoted many years of his life to the extensive breeding and systematic characterization of numerous generations of pea plants. He was the first to recognize a pattern of heritability, and dominant and recessive alleles. However, at that time Mendel was largely ignored by the scientific community and any molecular or biological knowledge supporting his hypothesis was lacking [1]. A century followed in which scientists progressively made discoveries, such as the heritable units being located on chromosomes inside the cell nucleus and a series of landmark studies from which ultimately the molecular structure of the molecule containing these units, deoxyribonucleic acid (DNA), was unraveled [2]. DNA is constructed of a sugar backbone and phosphate groups that all carry one out of four nitrogen-containing nucleobases: adenine (A), cytosine (C), guanine (G) and thymine (T). Adenine and thymine on the one hand and cytosine and guanine on the other are complementary bases and can form a covalent bond (base pair). Two complementary strands of DNA bind together and form the characteristic double-helix structure. The sequence of bases contained in the DNA is what makes up the ‘genetic code’. The discovery of the molecular structure of DNA and the genetic code was the start of a series of decades with other major discoveries in the field of (human) genetics. One of the major findings was that base-pair sequences on the DNA are transcribed into similar, single-stranded messenger molecules named messenger ribonucleic acid (mRNA) [3]. Triplets of mRNA bases are then translated into sequences of amino acids that are the building blocks of proteins, the active biological components in cells and biological systems [4]. The transcription and translation of the coding sequences in the DNA comprise the central dogma of molecular biology. Moreover, in the 1980s variation in the genetic code was revealed in the first maps of the human genome [5], and this variation was found to impact RNA, proteins and eventually biological systems and processes involved in health and disease.

1.2 Linking variation in the genetic code to disease

In the final two decades of the 20th century and the beginning of the 21st century, genetics of human disease mainly revolved around the discovery of highly penetrant variants with a large effect in so-called Mendelian diseases (e.g. Huntington's disease or cystic fibrosis); often rare diseases with a 'simple' genetic background where a mutation in one gene is causal (Figure 1). Linkage analysis, in which the segregation of a disease with genomic regions of interest was analyzed within families, was used for many Mendelian diseases to pinpoint a small region of the genome containing a causal variant in or near genes of interest [6]. In the same period, the Human Genome Project successfully completed the sequencing of the full human reference genome (> 3 billion base pairs) and made the entire genetic code available for the scientific community [7]. This project revealed an unprecedented amount of variation in the human genome. Among different types of genetic variants, the most frequently occurring (> 10 million in the genome) are single-nucleotide polymorphisms (SNPs). SNPs are single base pair differences that occur throughout the genome, spaced on average by a few hundred base pairs. The minor alleles of SNPs, the base pair configurations that are least present in the population, are common and have frequencies > 1% (minor allele frequency, MAF). Because of their high frequency in the population, effects of SNPs on disease are modest to low and they have incomplete penetrance (Figure 1). Therefore, linkage analyses turned out to be unsuitable to study associations between SNPs and diseases [6]. Elaborating on the efforts of the Human Genome project, the International HapMap project genotyped individuals from populations across the world and provided insight into the differences in genetic variation between populations [8]. The Human Genome and HapMap projects provided comprehensive maps of human genetic variation and allowed to define (population-specific) regions in the genome where alleles at neighboring polymorphisms were more frequently observed together than expected by chance (haplotypes) based on the principle of linkage disequilibrium (LD). Consequently, the need for a novel technique to genotype common genetic variants resulted in the design of SNP arrays: small chips containing thousands to millions of probes that bind complementary genomic DNA and emit a fluorescent signal of which the intensity reflects differences in binding efficiency of a DNA fragment containing a SNP, from which the SNP's genotype can subsequently be derived [9]. SNP

arrays allowed researchers to genotype large groups of unrelated individuals and obtain rich datasets capturing common genetic variation in haplotype blocks scattered across the genome. Using datasets with fully sequenced genomes and information on haplotypes, imputation methods were developed to infer unknown genotypes of variants in LD with SNPs genotyped on the array, adding information for additional millions of SNPs to these datasets [10].

1.3 Genome-wide association studies

Large-scale genotyping experiments using SNP arrays allowed to not just study the association to disease of several candidate genes or regions of the genome, but to perform genome-wide analyses. To achieve this, a statistical framework called the genome-wide association study (GWAS) was developed. Briefly, GWAS aims to detect association between a genetic variant and a phenotype of interest by correlating SNP allele frequencies to measured differences in a quantitative phenotype, or by looking at allele frequency differences in a dichotomous phenotype with a group of affected (cases) and unaffected (controls) individuals [11]. Association tests are performed independently across millions of SNPs. Thus, a strict significance threshold was defined to in order to call a genetic variant associated to a phenotype (p-value $< 5 \times 10^{-8}$, the genome-wide significance threshold) [12]. Furthermore, comprehensive quality control of data is required to remove SNPs that were not reliably measured on SNP arrays, and to elucidate risk of confounding due to e.g. allele frequency differences caused by population structure or relatedness between individuals [13]. Because of LD, an association is usually observed at a genomic region (a locus) in which associated variants possibly tag a causal variant. GWAS thus only highlights a small genomic region without pinpointing the exact causal variant. Nonetheless, the major advantage of GWAS over linkage studies is that it allows to study complex diseases with low effect variants in a hypothesis-free framework where the entire genome is scanned for associations [11,14].

The first landmark GWAS was performed in 2007 and since then, GWASs have successfully discovered thousands of variants associated with hundreds of traits and diseases, currently totaling > 68,000 SNP-trait associations [15-18]. The enormous increase in discovered variants has been the result of boosted statistical power in GWAS due to larger sample sizes, improved coverage of the

genome through development of novel SNP arrays and reference panels for imputation, and the development of novel statistical approaches [17]. Another major factor that influences power in GWAS is the phenotype definition, or how heterogeneous the studied phenotype is [17,19]. Especially in diseases that show variability in phenotypic presentation and are difficult to measure, such as several psychiatric disorders, this poses challenges to the discovery of disease-associated loci. To date, for almost all diseases the total expected contribution of common genetic variants to the heritability (SNP-based heritability, h_S^2) is higher than what we can predict based on GWAS findings. There is thus still a lack of power in GWASs to detect all common variants associated with a disease (referred to as hiding heritability) [20], promoting to further expand GWAS efforts and tackle abovementioned limitations to further unravel genetic risk.

1.4 Biological implications of disease-associated loci

The discoveries of GWAS have allowed to generate novel hypotheses about underlying biological processes involved in diseases [17]. Nonetheless, it is not straightforward to translate GWAS findings to pathophysiological mechanisms. Besides the low effect sizes of variants identified through GWAS and challenges in pinpointing the causal variant in a locus due to LD, over 95% of the genome is non-coding and most genetic variants thus do not directly impact the sequence of a gene and the resulting structure of the protein [7]. Therefore, GWAS findings can usually not be subjected to classical functional follow-up experiments (e.g. using gene knock-out techniques in cell or animal models) as is the case for highly penetrant mutations in single-gene diseases. Nonetheless, one of the most accessible and frequently applied methods to provide biological context to GWAS results is through the use of gene set enrichment analysis methods which test the enrichment of disease-associated variants in selected sets of genes that are e.g. involved in a specific biological process or expressed in a particular tissue. Several powerful tools to perform such analyses have been developed over the past years [21]. Additionally, functional knowledge has been used to build gene sets which are collected in various curated databases, such as the Molecular Signatures Database (MSigDB), Kyoto Encyclopedia of Genes and Genomes (KEGG) or Gene Ontology (GO) [22-24].

1.5 Sharing of genetic variation between (disease) phenotypes

Mendel also observed that some of the phenotypic features of his pea plants always inherited together. For example, white flower color would always go together with colorless seed coats, implying that the same unit of heredity was responsible for these phenotypes [1]. This phenomenon, in which a single variant or gene affects two or more distinct phenotypic traits was termed pleiotropy in the early 20th century [25]. One of the most well-known examples of pleiotropy is observed in the disease phenylketonuria where mutations in the phenylalanine hydroxylase gene (*PAH*), that plays a role in the conversion of the amino acid phenylalanine into tyrosine, result in mental retardation, eczema and pigmentation deficits [26]. Pleiotropy is however not restricted to traits and diseases caused by single genes and highly penetrant mutations. The era after the Human Genome Project has revealed that many common genetic loci discovered in GWAS are associated with multiple traits, suggesting the sharing of common genetic and biological pathways between complex phenotypes [27] (Figure 1). For example, SNPs in the Calcium Voltage-Gated Channel Subunit Alpha1 C (*CACNA1C*) gene have been linked to clinically overlapping psychiatric diseases schizophrenia and bipolar disorder [28,29]. Pleiotropic effects of variants thus reflect observations of shared heritability and comorbidity from epidemiology and clinical practice [27]. Besides the sharing of single genetic loci or genes, GWAS has also allowed to capture shared effects at millions of genetic variants for multiple phenotypes and calculate genetic correlations using data from unrelated individuals, whereas before this was only possible by assessing the co-segregation of two diseases in sparsely available datasets of related individuals. Thus, concerning common variant data, genetic correlation is the proportion of (SNP-based) heritability that is shared between a pair of traits. Several methods to compute genetic correlation based on GWAS datasets have been developed over the past years [30,31]. Application of such methods to GWAS datasets of hundreds of phenotypes have mostly revealed sharing of common genetic variants within trait and disease groups (such as anthropometric traits or psychiatric diseases), but have also resulted in the identification of unexpected genetic links between phenotypes (such as BMI and major depression) [32]. In turn, GWAS datasets of diseases that share heritability can be combined in large cross-disorder association analysis to increase power for the detection of pleiotropic loci [28]. Overall, results of

cross-disorder studies can further define the genetic landscape of complex diseases and aid in the identification of disease associated variants and biological pathways which improve diagnostics and treatment possibilities.

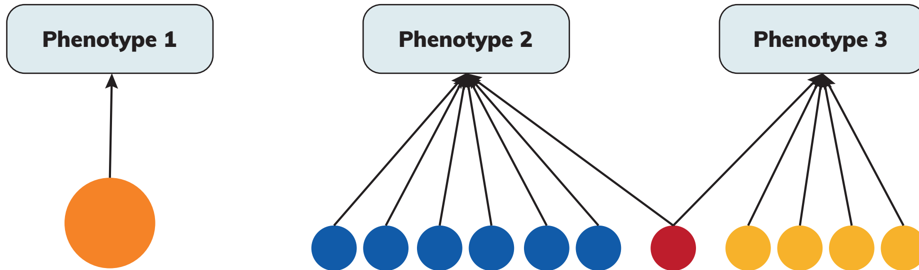


Figure 1. Mendelian- and polygenic phenotypes and pleiotropy. Circles indicate genetic variants, their size indicates their effect on a phenotype. Phenotype 1 is a 'simple' Mendelian phenotype where a variant with large effect (orange) is causal. Phenotype 2 and 3 are polygenic and are influenced by many common variants (SNPs) with low effect sizes. Besides phenotype-specific variants (blue and yellow), the SNP in red is pleiotropic as it affects both phenotype 2 and 3.

1.6 Psychiatric and neurological diseases studied in this thesis

Common genetic variation, GWAS, subsequent biological interpretation and shared genetic risk are major topics in this thesis that focuses on multiple complex psychiatric and neurological diseases which are briefly introduced in this section.

1.6.1 Schizophrenia

Schizophrenia is a heterogeneous psychiatric disorder that is characterized by positive symptoms (hallucinations, delusions), negative symptoms (social withdrawal, emotional flattening, anhedonia) and cognitive dysfunction [33]. It is a common disorder with an incidence of 15.2 per 100,000 persons per year and a lifetime prevalence of 1 in 100. Schizophrenia patients have increased mortality, mainly through suicide in early stages of the disease or cardiovascular disease [34]. The pathophysiological mechanisms of schizophrenia are unknown and there are no clear pathological hallmarks in the brain [35]. In contrast, a reduction of whole-brain volume has been reported in patients compared to healthy controls, and this reduction is more prominent in patients with a poor outcome [36]. Signs of neurodegeneration are however lacking [35]. Nonetheless, the implication of the dopaminergic system in schizophrenia is widely accepted

[37]. This is further supported by the fact that antipsychotic medication, the main and to date only effective treatment modality in schizophrenia, acts by blocking the dopaminergic system [38].

Genetic factors play an important role in the development of schizophrenia. Twin studies have estimated that the heritability is as high as 45%-80% [39,40]. A substantial part of this heritability is explained by SNPs (h_S^2 = 0.23) [41] and GWASs in schizophrenia have discovered genome-wide significant associations at over one hundred loci in large international collaborations [42,43]. Among the strongest associations are the major histocompatibility complex on chromosome 6 and the dopamine receptor D2 gene (*DRD2*) [42]. These insights into genetic risk and -architecture of schizophrenia allow to further study these findings in a biological context to generate novel hypotheses about disease mechanisms and pharmacological targets.

1.6.2 Stress- and trauma-related psychopathology

Exposure to severely stressful and traumatic events can result in the development of a range of psychiatric disorders, predominantly post-traumatic stress disorder (PTSD) and depression (or major depressive disorder, MDD). PTSD is characterized by a wide range of symptoms including intrusion, avoidance, negative effects on cognition and mood, aggression and disturbed sleep that are experienced for a prolonged period of time (> 1 month) [44,45]. Type, duration and number of traumatic events influence the severity and persistence of PTSD symptoms [46]. Depression has symptoms in common with PTSD, and is mainly characterized by changes in mood, interests and pleasure, cognition, and loss of appetite and weight [47]. The risk for MDD is especially well-established in relation to adverse events during childhood, with a twofold increased risk of developing the disease in individuals with a history of childhood trauma [48]. Particular professional groups, such as military personnel, are at increased risk of being exposed to trauma and to develop post-trauma psychopathology [49]. Additionally, there is a comorbidity of depression in almost half of the military PTSD cases [50]. The hypothalamic-pituitary-adrenal (HPA) axis, the main system involved in stress regulation, plays a key role in the development of stress-related psychiatric disease through deficient glucocorticoid signaling, which is likely further influenced by regulating factors including inflammatory cytokines, neurosteroids or biochemical substances

that mediate resilience [44,47]. The main treatment of PTSD and trauma-related depression comprises psychotherapy that focusses on processing of the trauma and reducing adverse effects. In MDD, pharmacotherapy with monoamine oxidase inhibitors also belongs to the regular treatment options. Limited understanding of PTSD has hampered the development of equally efficient pharmacological therapies [44,47].

Early candidate-gene studies identified associations of polymorphisms in several HPA-axis-related genes, such as *FKBP5*, with the development of psychopathology after trauma exposure [51,52]. Until recently, PTSD had not been studied in GWAS as extensively as other psychiatric diseases. The first GWAS effort in PTSD identified one genome-wide significant locus in the retinoid-related orphan receptor alpha gene (RORA) [53]. Consecutive GWASs have identified several additional loci, although often without successful replication of previously discovered loci. Recently, insight into the SNP-based heritability was provided by a large GWAS led by the Psychiatric Genomics Consortium (PGC) that showed PTSD has moderate SNP-based heritability in women (h_S^2 = 29%), while heritability was low and non-significant in men (h_S^2 = 7%) [54]. In contrast to PTSD, GWAS in MDD has successfully identified dozens of disease-associated loci in studies including up to half a million individuals and have established a moderate role for common genetic variants (h_S^2 = 8.7%) [55,56]. However, these studies were not specifically performed in trauma-exposed individuals.

1.6.3 Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is a severe neurodegenerative disease characterized by the progressive loss of upper- and lower motor neurons, resulting in increased weakening of spinal and bulbar muscles and respiratory failure in the final stage of the disease [57]. ALS has an incidence of 3.3 per 100,000 persons per year and a lifetime risk 1 in 400 with a median survival of 3-5 years after onset of the first symptoms [58,59]. While survival can be extended with a limited number of months through administration of riluzole, no effective treatment for the disease currently exists [60]. Besides involvement of the motor system, cognitive and behavioral symptoms are not rare in ALS with ~34% of patients fulfilling criteria for cognitive impairment and ~13% meeting diagnostic criteria for co-morbid frontotemporal dementia (FTD) [61].

Another widely observed symptom among ALS patients are muscle cramps, likely caused by imbalance in the excitability of motor neurons, which can result in pain [62].

The importance of genetics in ALS is reflected by the total heritability of the disease of 65% [63]. Around 10-20% of ALS cases from European descent have a family history of the disease [57], in which rare mutations in the genes *C9orf72*, *TARDBP*, *SOD1* and *FUS* play a major role [64]. However, mutations in these genes do not explain all variation in familial ALS cases, implying a role for common genetic variants. Therefore, it has been suggested that ALS has an oligogenic disease architecture in which both rare and common variants seem to play a role [65]. Furthermore, the majority of patients has a sporadic form of the disease which cannot be explained by rare familial mutations. GWASs have provided insight into the common genetic architecture of ALS with the most recent SNP-based heritability of ALS being estimated at 8.5% while several associated loci have been discovered and replicated [66].

1.6.4 Epilepsy

Epilepsy is a common neurological disorder characterized by the enduring predisposition to generate epileptic seizures, caused by abnormal and excessive synchronous neuronal activity in the brain [67,68]. One in 26 people develops epilepsy during their lifetime and the disease has an overall incidence of 67.7 per 100,000 persons per year, a lifetime prevalence of 7.6 per 1,000 and an increased mortality rate compared to the general population, making it one of the most common neurological disorders with a large burden of disease [69-71]. Two major subtypes of epilepsy can be differentiated by the origin of seizures: focal (seizures originate in localized or specific areas of the brain) and generalized (widespread occurrence of seizures across both hemispheres) and both seizure types can include motor involvement [68]. The first choice of treatment comprises a variety of anti-seizure drugs (ASDs), but these are only effective in two thirds of the patients and around 80% develops side-effects that reduce their quality of life. Surgery is a second treatment option but is only possible depending on the location in the brain where seizures originate [68].

Although a variety of acquired causes for the disease exist (e.g. traumatic brain injury or infectious disease), genetic predisposition also plays a role. Rare

variants and copy number variants (CNVs) have been identified in familial epilepsy, but common genetic variants also play an important role (h_S^2 estimated at 26%) and an increasing number of genome-wide significant loci have been identified through GWAS [72-74]. Recent GWAS has shown a greater involvement of common genetic variants in generalized epilepsy compared to focal epilepsy [74].

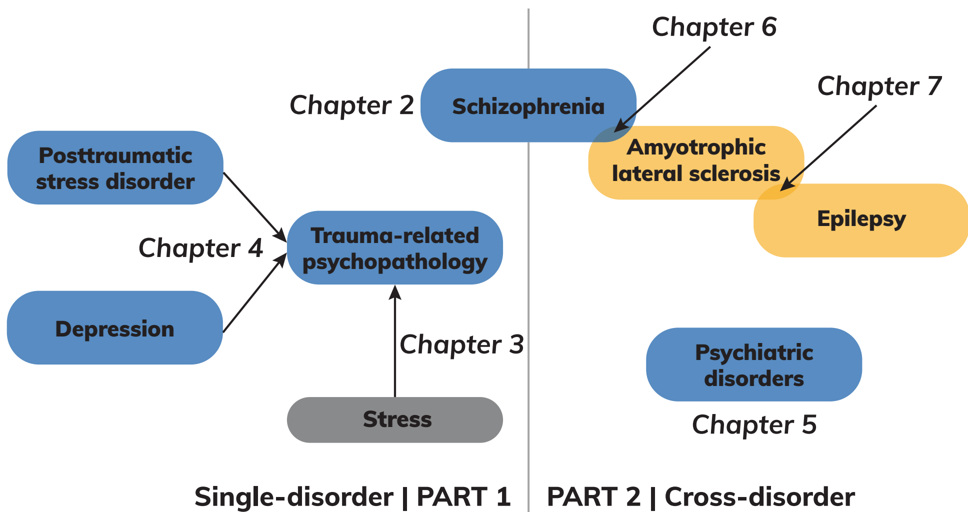


Figure 2. Schematic overview of psychiatric and neurological phenotypes covered in this thesis and the context in which they are studied.

1.7 Contents and aims of this thesis

This thesis consists of two parts containing single- (***PART 1***) and cross-disorder (***PART 2***) genome-wide analyses (Figure 2):

PART 1 focuses on psychiatric phenotypes, specifically schizophrenia and trauma-related psychopathology (mainly PTSD and depression). Lack of knowledge about the biological pathways involved in schizophrenia hampers the development of novel treatments. Using results from large schizophrenia GWASs, we aim to investigate which biological gene sets and pathways are associated with the disease in a comprehensive analysis in ***chapter 2***. Like schizophrenia, post-trauma psychopathology has a heterogeneous clinical

presentation. GWAS studies on PTSD are in a relatively early stage and are challenged by a large heterogeneity in cohort and study designs. ***Chapter 3*** describes a multivariate GWAS analysis on quantitative phenotypes related to stress and psychopathology in a prospective longitudinal cohort of Dutch military personnel. In the same military cohort, the effect of common genetic risk for PTSD and MDD on the development of these disorders is assessed in ***chapter 4***, taking into account the interaction with trauma experienced during childhood and during deployment.

PART 2 contains cross-disorder studies. In ***chapter 5***, we outline the main clinical domains in psychiatry which are impacted by these studies and explain important analysis techniques. ***Chapter 6*** describes the largest study on shared common genetic risk between ALS and schizophrenia. As cognitive impairment plays a role in both diseases, we hypothesize that this overlap in clinical features might be reflected at the genetic level. Furthermore, the role of motor neuron excitability in ALS, leading to muscle cramps, has been challenged by several hypotheses. In ***chapter 7***, cross-disorder genome-wide analyses with epilepsy, a disease in which neuronal excitability plays a major role, are performed to test the possibility of shared genetic risk.

Overall, these studies have the mutual goal to discover novel risk loci associated with disease; gain insight into shared heritability, genetic architecture and the genetic landscape in which these diseases are present; and provide novel hypotheses about disease biology and etiology.

1.8 References

1. Fairbanks DJ, Rytting B. Mendelian controversies: a botanical and historical review. Am J Bot 2001;88(5):737–52.
2. Watson J, Crick F. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953;171(4356):737–8.
3. Brenner S, Jacob F, Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 1961;190:576–81.
4. Nirenberg MW, Matthaei JH. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 1961;47(10):1588–602.
5. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, et al. A genetic linkage map of the human genome. Cell 1987;51(2):319–37.
6. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 2003;33(3s):228–37.
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860–921.
8. The International HapMap Consortium. A haplotype map of the human genome. Nature 2005;437(7063):1299–320.
9. LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res 2009;37(13):4181–93.
10. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010;11(7):499–511.
11. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6(2):95–108.
12. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 2008;32(4):381–5.
13. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc 2010;5(9):1564–73.
14. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med 2009;360(17):1759–68.
15. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661–78.
16. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012;90(1):7–24.
17. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017;101(1):5–22.
18. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014;42(D1):D1001–6.
19. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PLoS One 2013;8(10):e76295.
20. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. Nat Rev Genet 2014;15(11):765–76.
21. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nat Rev Genet 2016;17(6):353–64.
22. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res 2015;43(Database issue):D1049–56.
23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30.

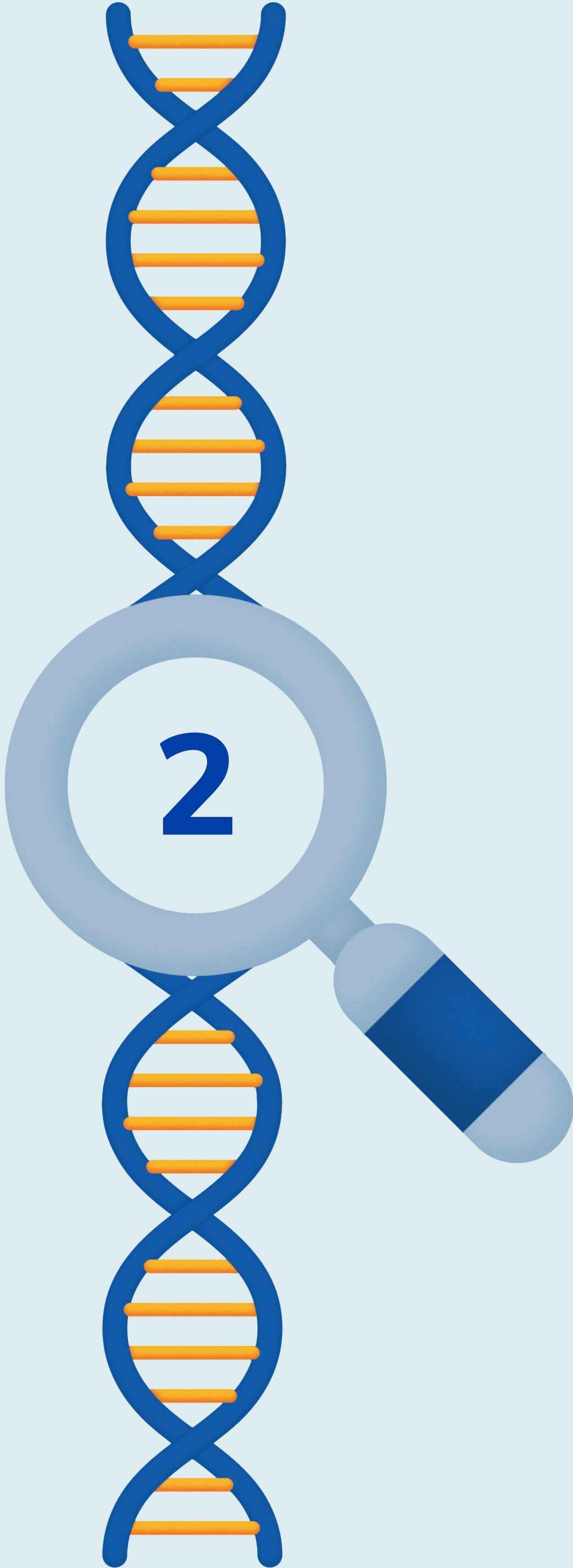
24. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27(12):1739–40.
25. Stearns FW. One hundred years of pleiotropy: a retrospective. Genetics 2010;186(3):767–73.
26. Blau N. Genetics of Phenylketonuria: Then and Now. Hum Mutat 2016;37(6):508–15.
27. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 2013;14(7):483–95.
28. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013;381(9875):1371–9.
29. Laursen TM, Agerbo E, Pedersen CB. Bipolar disorder, schizoaffective disorder, and schizophrenia overlap: a new comorbidity index. J Clin Psychiatry 2009;70(10):1432–8.
30. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.
31. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 2012;28(19):2540–2.
32. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science 2018;360(6395):eaap8757.
33. Kahn RS, Sommer IE, Murray RM, Meyer-Lindenberg A, Weinberger DR, Cannon TD, et al. Schizophrenia. Nat Rev Dis Primers 2015;:15067.
34. McGrath J, Saha S, Chant D, Welham J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol Rev 2008;30(1):67–76.
35. Harrison PJ, Weinberger DR. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Mol Psychiatry 2005;10(1):40–68–image5.
36. Haijma SV, Van Haren N, Cahn W, Koolschijn PCMP, Hulshoff Pol HE, Kahn RS. Brain Volumes in Schizophrenia: A Meta-Analysis in Over 18 000 Subjects. Schizophr Bull 2012;39(5):1129–38.
37. Howes OD, Williams M, Ibrahim K, Leung G, Egerton A, McGuire PK, et al. Midbrain dopamine function in schizophrenia and depression: a post-mortem and positron emission tomographic imaging study. Brain 2013;136(Pt 11):3242–51.
38. Kane JM, Correll CU. Past and present progress in the pharmacologic treatment of schizophrenia. J Clin Psychiatry 2010;71(9):1115–24.
39. Lichtenstein P, Björk C, Hultman CM, Scolnick E, Sklar P, Sullivan PF. Recurrence risks for schizophrenia in a Swedish national cohort. Psychol Med 2006;36:1417–25.
40. Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. Nat Genet 2017;6:111.
41. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013;45(9):984–94.
42. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511(7510):421–7.
43. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 2018;199:441.
44. Yehuda R, Hoge CW, McFarlane AC, Vermetten E, Lanius RA, Nievergelt CM, et al. Post-traumatic stress disorder. Nat Rev Dis Primers 2015;1:15057.
45. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5 ed. Washington, D.C.: 2013.
46. Karam EG, Friedman MJ, Hill ED, Kessler RC, McLaughlin KA, Petukhova M, et al. Cumulative traumas and risk thresholds: 12-month PTSD in the World Mental Health (WMH) surveys. Depress Anxiety 2014;31(2):130–42.
47. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. Nat Rev Dis Primers 2016;2:16065.

48. Hovens JGFM, Giltay EJ, Wiersma JE, Spinhoven P, Penninx BWJH, Zitman FG. Impact of childhood life events and trauma on the course of depressive and anxiety disorders. Acta Psychiatr Scand 2012;126(3):198–207.
49. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N Engl J Med 2004;351(1):13–22.
50. Walter KH, Levine JA, Highfill-McRoy RM, Navarro M, Thomsen CJ. Prevalence of Posttraumatic Stress Disorder and Psychological Comorbidities Among U.S. Active Duty Service Members, 2006-2013. J Trauma Stress 2018;61(3):22.
51. Binder EB, Bradley RG, Liu W, Epstein MP, Deveau TC, Mercer KB, et al. Association of FKBP5 Polymorphisms and Childhood Abuse With Risk of Posttraumatic Stress Disorder Symptoms in Adults. JAMA 2008;299(11):1291–305.
52. Rao S, Yao Y, Ryan J, Li T, Wang D, Zheng C, et al. Common variants in FKBP5 gene and major depressive disorder (MDD) susceptibility: a comprehensive meta-analysis. Sci Rep 2016;6(1):32687.
53. Logue MW, Baldwin C, Guffanti G, Melista E, Wolf EJ, Reardon AF, et al. A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. Mol Psychiatry 2013;18(8):937–42.
54. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. Mol Psychiatry 2018;23(3):666–73.
55. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet 2016;48(9):1031–6.
56. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018;50(5):668–81.
57. Hardiman O, Al-Chalabi A, Chiò A, Corr EM, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. Nat Rev Dis Primers 2017;3:17071.
58. Hardiman O, van den Berg LH, Kiernan MC. Clinical diagnosis and management of amyotrophic lateral sclerosis. Nat Rev Neurol 2011;7(11):639–49.
59. Alonso A, Logroscino G, Jick SS, Hernán MA. Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study. Eur J Neurol 2009;16(6):745–51.
60. Beghi E, Chiò A, Couratier P, Esteban J, Hardiman O, Logroscino G, et al. The epidemiology and treatment of ALS: focus on the heterogeneity of the disease and critical appraisal of therapeutic trials. Amyotroph Lateral Scler 2011;12(1):1–10.
61. Phukan J, Elamin M, Bede P, Jordan N, Gallagher L, Byrne S, et al. The syndrome of cognitive impairment in amyotrophic lateral sclerosis: a population-based study. J Neurol Neurosurg Psychiatry 2012;83(1):102–8.
62. Stephens HE, Joyce NC, Oskarsson B. National Study of Muscle Cramps in ALS in the USA. Amyotroph Lateral Scler Frontotemporal Degener 2017;18(1-2):32–6.
63. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry 2010;81(12):1324–6.
64. Chiò A, Battistini S, Calvo A, Caponnetto C, Conforti FL, Corbo M, et al. Genetic counselling in ALS: facts, uncertainties and clinical suggestions. J Neurol Neurosurg Psychiatry 2014;85(5):478–85.
65. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci 2014;17(1):17–23.
66. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet 2016;48(9):1043–8.
67. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al. ILAE official report: a practical clinical definition of epilepsy. Epilepsia 2014;55(4):475–82.

68. Devinsky O, Vezzani A, O'Brien TJ, Jette N, Scheffer IE, de Curtis M, et al. Epilepsy. Nat Rev Dis Primers 2018;4:18024.
69. Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon C-S, Dykeman J, et al. Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. Neurology 2017;88(3):296–303.
70. Thurman DJ, Logroscino G, Beghi E, Hauser WA, Hesdorffer DC, Newton CR, et al. The burden of premature mortality of epilepsy in high-income countries: A systematic review from the Mortality Task Force of the International League Against Epilepsy. Epilepsia 2017;58(1):17–26.
71. Hesdorffer DC, Logroscino G, Benn EKT, Katri N, Cascino G, Hauser WA. Estimating risk for developing epilepsy: a population-based study in Rochester, Minnesota. Neurology 2011;76(1):23–7.
72. Speed D, O'Brien TJ, Palotie A, Shkura K, Marson AG, Balding DJ, et al. Describing the genetic architecture of epilepsy through heritability analysis. Brain 2014;137(Pt 10):2680–9.
73. The International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol 2014;13(9):893–903.
74. The International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun 2018;9(1):5269.

Part 1

Single-disorder studies


2

Chapter 2

Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling

Dick Schijven, Daniel Kofink, Vinicius Tragante, Marloes Verkerke, Sara L. Pulit, René S. Kahn, Jan H. Veldink, Christiaan H. Vinkers, Marco P. Boks, Jurjen J. Luykx

Published in *Schizophrenia Research*, 199:195-202 (2018)

Abstract

Large-scale genome-wide association studies (GWAS) have implicated many low-penetrance loci in schizophrenia. However, its pathological mechanisms are poorly understood, which in turn hampers the development of novel pharmacological treatments. Pathway and gene set analyses carry the potential to generate hypotheses about disease mechanisms and have provided biological context to genome-wide data of schizophrenia. We aimed to examine which biological processes are likely candidates to underlie schizophrenia by integrating novel and powerful pathway analysis tools using data from the largest Psychiatric Genomics Consortium schizophrenia GWAS (N = 79,845) and the most recent 2018 schizophrenia GWAS (N = 105,318). By applying a primary unbiased analysis (Multi-marker Analysis of GenoMic Annotation; MAGMA) to weigh the role of biological processes from the Molecular Signatures Database (MSigDB), we identified enrichment of common variants in synaptic plasticity and neuron differentiation gene sets. We supported these findings using MAGMA, Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) and Interval Enrichment Analysis (INRICH) on detailed synaptic signaling pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and found enrichment in mainly the dopaminergic and cholinergic synapses. Moreover, shared genes involved in these neurotransmitter systems had a large contribution to the observed enrichment, protein products of top genes in these pathways showed more direct and indirect interactions than expected by chance, and expression profiles of these genes were largely similar among brain tissues. In conclusion, we provide strong and consistent genetics and protein-interaction informed evidence for the role of postsynaptic signaling processes in schizophrenia, opening avenues for future translational and psychopharmacological studies.

2.1 Introduction

Although post-mortem studies, imaging and human genetic studies have contributed to theories about pathophysiological mechanisms in schizophrenia, the underlying molecular processes have not been fully elucidated. This knowledge gap hampers the development of novel pharmacological treatments. Genetic studies provide a valuable resource to investigate the mechanisms that are likely at play in schizophrenia. Schizophrenia is highly heritable (h^2 estimates ranging from 45 to 80%) and polygenic [1,2]. The two largest genome-wide association studies (GWAS) have identified 108 [3] and 145 independent associated risk loci [4].

Pathway and gene set enrichment analysis methods are widely used to provide biological context to the results of genetic association studies by testing whether biologically relevant pathways or sets of genes are enriched for genetic variants associated with a phenotype [5]. These analyses have been widely applied to schizophrenia, providing evidence for the involvement of synaptic and immune-related processes [6,7] and insight into possible new drug targets [8]. Such findings are supported by pathway analyses in combined psychiatric disorders (schizophrenia, depression and bipolar disorder), revealing enrichment of genetic variants in neuronal, immune and histone pathways [9]. Involvement of calcium signaling and ion channels in schizophrenia has been reported in a gene set analysis paper combining GWAS data with post-mortem brain gene expression data [10]. Importantly, none of the abovementioned pathway analysis studies has used the full dataset reported in the latest Psychiatric Genomics Consortium schizophrenia GWAS [3]. Moreover, several novel or widely used pathway analysis tools have not yet been applied to this schizophrenia GWAS. These tools constitute fast and powerful approaches to test gene set enrichment, despite variability in their correcting for confounding factors that may increase the type 1 error rate [5]. Additionally, tools and databases aimed at the integration of GWAS data with gene expression and protein-protein interaction data allow to further explore the biological impact of common variants associated with schizophrenia [11,12].

Aiming to comprehensively investigate the possible biological processes underlying schizophrenia, we set out to apply gene set and pathway enrichment

analysis methods to the 2014 schizophrenia GWAS [3] and additionally integrate the results of these analyses with data on protein-protein interactions and tissue-specific gene expression (Figure 1). We then validated our main findings using the most recent 2018 schizophrenia GWAS [4]. We thus elucidate the involvement of neuron differentiation and synaptic plasticity in schizophrenia and reveal an accumulation of variants in post-synaptic signaling cascades. The analyses moreover enable a more nuanced understanding of the several actionable classes of neurotransmitters implicated in the disease.

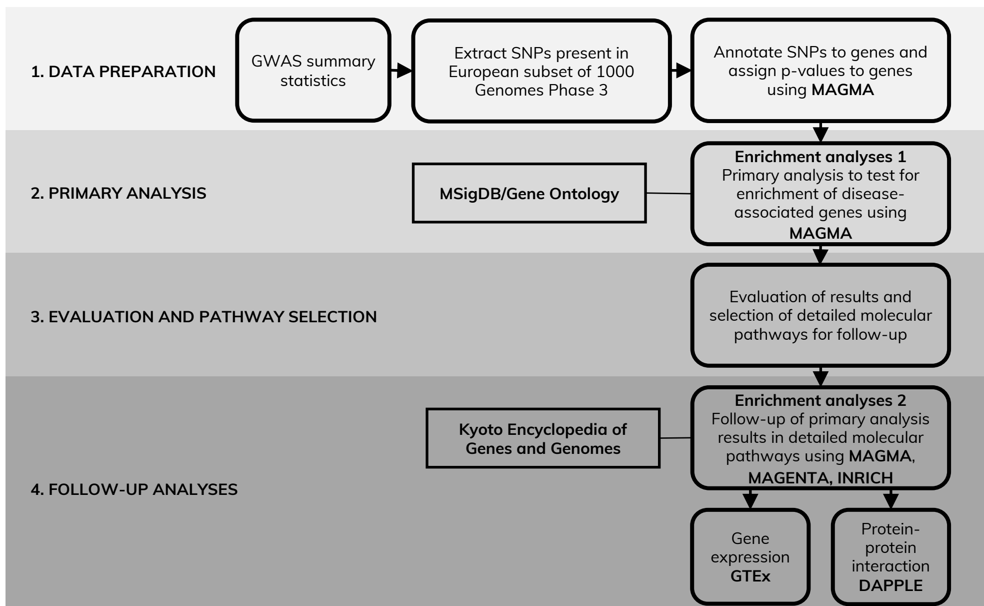


Figure 1. Overview of pathway analysis pipeline. Our analysis pipeline consisted of four stages: 1) Data preparation; 2) Primary gene set enrichment analysis on MSigDB gene ontology (GO) biological processes; 3) Evaluation of primary results and selection of detailed pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG); 4) Follow-up analyses on these detailed molecular pathways to further investigate involvement of biological processes found in the primary analysis.

2.2 Materials and methods

2.2.1 *Input data and analysis overview*

We used summary-level results from the largest and publicly available Psychiatric Genomics Consortium GWAS in schizophrenia (www.med.unc.edu/pgc/results-and-downloads; downloaded on 10 May 2017) [3]. This GWAS was performed in 34,241 schizophrenia cases and 45,604 healthy controls and

association results for 9.4 million single-nucleotide polymorphisms (SNPs) were reported in the summary-level data. As detailed below (also see Figure 1), using Multi-marker Analysis of GenoMic Annotation (MAGMA) [13] we successively (A) mapped SNPs to genes, (B) calculated gene p-values based on GWAS SNP p-values, (C) performed a primary gene set enrichment analysis using gene ontology (GO) terms, and (D) tested the robustness of these findings in detailed molecular pathways derived from KEGG. We then validated these findings using summary statistics derived from the latest 2018 schizophrenia GWAS [4] (downloaded on 1 March 2018). Finally, we further investigated the results of the analysis on KEGG pathways using Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) and Interval Enrichment Analysis (INRICH) [14,15], applied in silico protein-protein interaction (PPI) analysis using Disease Association Protein-Protein Link Evaluator (DAPPLE) [11], and assessed tissue-specific expression using data from the Genotype-Tissue Expression (GTEx) project [12].

2.2.2 *Mapping SNPs to genes and assigning p-values to genes*

SNPs present in the European subset of the 1000 Genomes Phase 3 dataset were extracted from the GWAS summary-level results [16]. Using MAGMA v1.06, we mapped SNPs to corresponding genes, extending gene footprints by an additional 20 kilobase (kb) up- and downstream, as a large proportion of regulatory elements involved in gene expression regulation is likely to be captured by including this region [17]. We then applied a gene analysis to obtain a p-value for each gene to which at least one SNP was mapped. The p-value of a gene was based on the mean association statistic of SNPs contained in that gene. Linkage disequilibrium (LD) between SNPs included in the gene analysis was estimated from the European subset of 1000 Genomes Phase 3, which corresponds best with the ancestry of samples in the schizophrenia GWAS [3].

2.2.3 *Gene Ontology (GO) gene set enrichment analysis*

We downloaded 4,436 GO biological process gene sets from the Molecular Signatures Database (MSigDB, release 6.0, April 2017) [18,19]. MSigDB merged original GO terms with high similarity (Jaccard's coefficient > 0.85) into one gene set and the number of genes per gene set was limited to $10 < n_{genes} < 2000$. We applied a competitive gene set analysis, which tests whether the genes in each gene set are more strongly associated with

schizophrenia than genes outside the gene set, using gene p-values from the mapping step. The significance threshold (at $\alpha = 0.05$) was adjusted for multiple GO gene sets tested using a permutation procedure implemented in MAGMA (10,000 permutations, $p = 4.23 \times 10^{-6}$). A correction for confounding factors (gene size, gene density and minor allele count) was applied. We additionally tested the robustness of our analysis by implementing several sensitivity analyses: 1) excluding the highly associated MHC region on chromosome 6 (chr6:25,500,000–33,500,000, human genome assembly GRCh37/hg19), 2) excluding the X-chromosome from the analysis, and 3) applying strict filtering for gene set size ($10 < n_{genes} \leq 200$) as larger gene sets are sometimes perceived as too broad or uninterpretable.

2.2.4 Follow-up using KEGG pathways

To further test the results of our primary GO gene set analysis, we used pathways representing synaptic signaling processes from the Kyoto Encyclopedia of Genes and Genomes (KEGG, www.kegg.jp, downloaded on 3 January 2017) [20] as input for MAGMA: glutamatergic synapse (hsa04724, 114 genes), cholinergic synapse (hsa04725, 111 genes), serotonergic synapse (hsa04726, 113 genes), GABAergic synapse (hsa04727, 88 genes), dopaminergic synapse (hsa04728, 130 genes), long-term potentiation (hsa04720, 67 genes), and long-term depression (hsa04730, 60 genes). We then ran a similar competitive gene set analysis on these pathways (multiple-testing corrected $p = 8.6 \times 10^{-3}$). For each significantly associated pathway, we mapped top associated genes (gene $p < 1.45 \times 10^{-4}$: 0.05/344 unique genes included in the enrichment analysis of the selected KEGG pathways) to pathway components using *pathview* in R version 3.3.3 (www.r-project.org) [21]. Because many genes overlapped between two or more of the tested KEGG pathways (Supplementary Figure 1), we conditioned the enrichment specific to each KEGG pathway on the association signal of genes shared with other tested KEGG pathways (built-in function in MAGMA). Additionally, we tested enrichment of sets of unique genes per KEGG pathway and a gene set containing all shared genes.

To support the results of the MAGMA analysis, we tested enrichment in the same KEGG pathways using MAGENTA and INRICH. MAGENTA calculated gene p-values based on the lowest SNP p-value in a gene (extended 20 kb up-

and downstream) while correcting for confounding factors, such as gene length and gene overlap. It then assessed whether a gene set was enriched with low gene p-values at the 95th percentile cut-off (based on all gene p-values in the genome) compared to randomly sampled gene sets of similar size [15]. Enrichment p-values were corrected for multiple testing using false discovery rate (FDR < 0.05). INRICH tested whether intervals around associated variants overlap more often with genes in a gene set than could be expected by chance [14]. Intervals were calculated around genome-wide significant SNPs ($p < 5 \times 10^{-8}$) using the clump function in PLINK v1.90b3z [22]. Only SNPs with a LD $R^2 > 0.5$ and an association p-value < 0.05 were included in an interval, resulting in 114 non-overlapping intervals. Target gene regions were extended with 20 kb up- and downstream. Enrichment of KEGG pathways was tested using a permutation procedure where the enrichment of a gene set was tested against the normal distribution of enrichment in background gene sets. Enrichment p-values were corrected for multiple testing by a bootstrap method in INRICH.

2.2.5 Protein-protein interaction analysis

To increase insight into the biological impact of the pathways implicated in schizophrenia by our analyses, we performed in silico PPI analysis using DAPPLE v0.17 [11]. We used DAPPLE to test whether proteins in a network comprised of 22 unique top genes from the KEGG dopaminergic, cholinergic or long-term potentiation pathways (gene $p < 1.45 \times 10^{-4}$) show more direct and indirect interactions with each other and with other proteins than expected by chance.

2.2.6 Gene expression analysis

Median relative transcript abundances (transcripts per million, TPM) among 13 brain tissues and 40 non-brain tissues were obtained from the GTEx database (v7, RNASeQCv1.8.8, downloaded on 1 November 2017 from www.gtexportal.org/home/datasets) [12]. Per enriched KEGG pathway, we extracted TPM levels of top genes (gene $p < 1.45 \times 10^{-4}$) and visualized the median relative abundance of transcripts in heat maps for each enriched pathway and for one set of genes that were shared between two or more of the enriched KEGG pathways using the R-package *gplots* [23]. Genes with high similarities in transcript abundance, as well as tissues with similar transcript

abundance, were clustered. Additionally, we used median gene expression level in brain (all brain tissues) as a covariate in the MAGMA competitive gene set analysis of two significant GO gene sets and the tested KEGG pathways to assess whether enrichment was dependent on brain gene expression.

2.3 Results

Primary gene set enrichment analysis in MAGMA (in 4,436 MSigDB GO biological processes) identified two gene sets that were enriched for schizophrenia-associated SNPs (Figure 2A and Supplementary Table 1): 'Regulation of Synaptic Plasticity' ($p = 6.52 \times 10^{-7}$) and 'Regulation of Neuron Differentiation' ($p = 1.00 \times 10^{-6}$), which are specialized GO terms of synaptic signaling and neuron differentiation (Supplementary Figure 2). Excluding the extended MHC region or the X-chromosome did not affect the outcome of this analysis. When we applied strict filtering for gene set size, we found that the 'Regulation of neuron differentiation' gene set (containing 554 genes) was no longer significant and only 'Regulation of Synaptic Plasticity' remained significant (Supplementary Figure 3). There was extensive coherence between the significant GO gene sets (54 shared genes, Supplementary Figure 1B), confirmed by an attenuation of enrichment when the analysis was conditioned on these shared genes (Supplementary Figure 4A).

To gain a more nuanced understanding of molecular synaptic pathways enriched for schizophrenia-associated variants, we tested enrichment of SNPs in KEGG pathways representing synaptic signaling. We found significant enrichment in pathways representing the dopaminergic synapse ($p = 1.62 \times 10^{-4}$), cholinergic synapse ($p = 1.51 \times 10^{-3}$) and long-term potentiation ($p = 4.00 \times 10^{-3}$) (Figure 2B, Supplementary Table 2). For each significant pathway, we mapped top genes (from the MAGMA gene analysis step, see Supplementary Table 3) on components within these KEGG pathways (Figure 3A, Supplementary Figure 5). SNP enrichment was mostly restricted to trans-membrane and postsynaptic components in the cholinergic and dopaminergic synapses. The long-term potentiation pathway only included post-synaptic components. We detected strong enrichment in signaling through extracellular signal-regulated kinase (ERK) and cAMP response element-binding protein (CREB), phospholipase C (PLC) and the inositol trisphosphate receptor (IP_3R),

and protein kinase B (PKB/Akt). These cascades converge on mechanisms involved in synaptic growth regulation and synaptic plasticity. Furthermore, voltage-gated calcium channels, glutamatergic NMDA and AMPA receptors, the dopamine D2 receptor (DRD2) and the muscarinic acetylcholine receptor M4 were highly enriched. A proportion of genes was shared between the tested KEGG pathways and enriched GO gene sets (Supplementary Figure 1B). Conditioning the analysis of the two significant GO gene sets on these shared genes revealed a slight effect of these genes on the enrichment signal (Supplementary Figure 4A). When we conditioned the analysis of each KEGG pathway on shared genes with other tested KEGG pathways, all enrichment p-values increased to non-significance ($p > 8.6 \times 10^{-3}$), indicating a considerable contribution of shared genes to the enrichment signal (Supplementary Figure 4B). Shared genes between KEGG pathways that were tested as a single gene set were however not enriched after multiple testing correction in a combined analysis with sets of unique genes per pathway ($p = 9.4 \times 10^{-3}$) (Supplementary Figure 6).

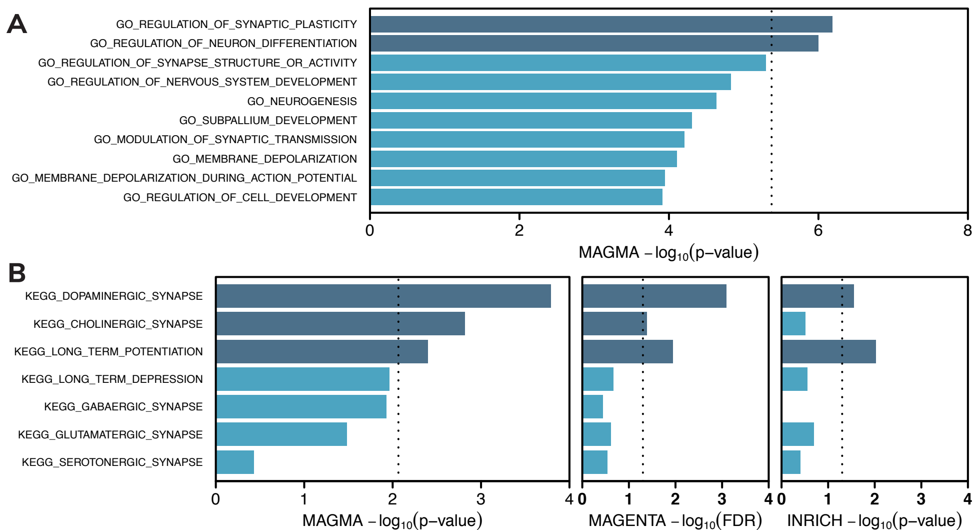


Figure 2. Results of enrichment analysis on MSigDB biological processes and KEGG pathways. Reported p-values or FDR are $-\log_{10}$ converted. Significance thresholds are different per analysis depending on the used method. Significant enrichment is indicated with dark blue. **A)** Top ten enriched MSigDB gene sets (significance threshold $p = 4.23 \times 10^{-6}$). **B)** Enrichment in KEGG synaptic signaling pathways tested using MAGMA (left panel, significance threshold $p = 8.6 \times 10^{-3}$), MAGENTA (middle panel, significance threshold FDR = 0.05) and INRICH (right panel, significance threshold p = 0.05).

Protein products of 22 unique top genes from enriched KEGG pathways showed more direct interactions than expected by chance ($p = 1.4 \times 10^{-2}$) and more indirect connections to other proteins ($p = 4.0 \times 10^{-3}$), but not direct connections to other proteins ($p = 0.81$) (Figure 3B,C). Gene expression analysis revealed no large differences in transcript abundance for top genes among brain tissues, although relative expression seems higher in cerebellum and cerebellar hemisphere compared to other brain tissues (Supplementary Figure 7), and between brain tissues and other tissues (Supplementary Figure 8). When we applied median gene expression levels in the brain as a covariate in the MAGMA competitive gene set analysis of GO gene sets and KEGG pathways, we observed a slight attenuation of the enrichment signal although all enriched GO gene sets and KEGG pathways remained significant (Supplementary Figure 9).
Using MAGENTA and INRICH with the same data and gene sets, we confirmed significant enrichment of the dopaminergic synapse (MAGENTA corrected FDR = 8.00×10^{-4}, INRICH corrected p-value = 2.82×10^{-2}) and long-term potentiation (MAGENTA corrected FDR = 1.14×10^{-2}, INRICH corrected p-value = 9.6×10^{-3}). Enrichment for the cholinergic synapse was only confirmed in MAGENTA (corrected FDR = 4.12×10^{-2}) (Figure 2B). When we applied our analysis pipeline to the data of the 2018 schizophrenia GWAS meta-analysis [4], we confirmed enrichment of neurodevelopmental processes and the dopaminergic synapse, while results again pointed towards enrichment of the cholinergic synapse. In addition, we also found enrichment of the GABAergic synapse in the MAGMA analysis. Only enrichment in the dopaminergic synapse was confirmed by all three gene set analysis tools (Supplementary Figure 10).

2.4 Discussion

By implementing complementary gene set enrichment analysis tools (MAGMA, MAGENTA and INRICH), annotations from biological databases (MSigDB/GO and KEGG), protein-protein interaction data, and tissue-specific gene expression into a comprehensive analysis, we aimed to elucidate biological processes underlying schizophrenia. Thus, we first detected enrichment of schizophrenia-associated SNPs in synaptic plasticity and neuronal differentiation processes, which is in line with biological hypotheses introduced previously [7,9,10]. We did however not replicate significant enrichment of immune and histone pathways (Supplementary Figure 11), which have also

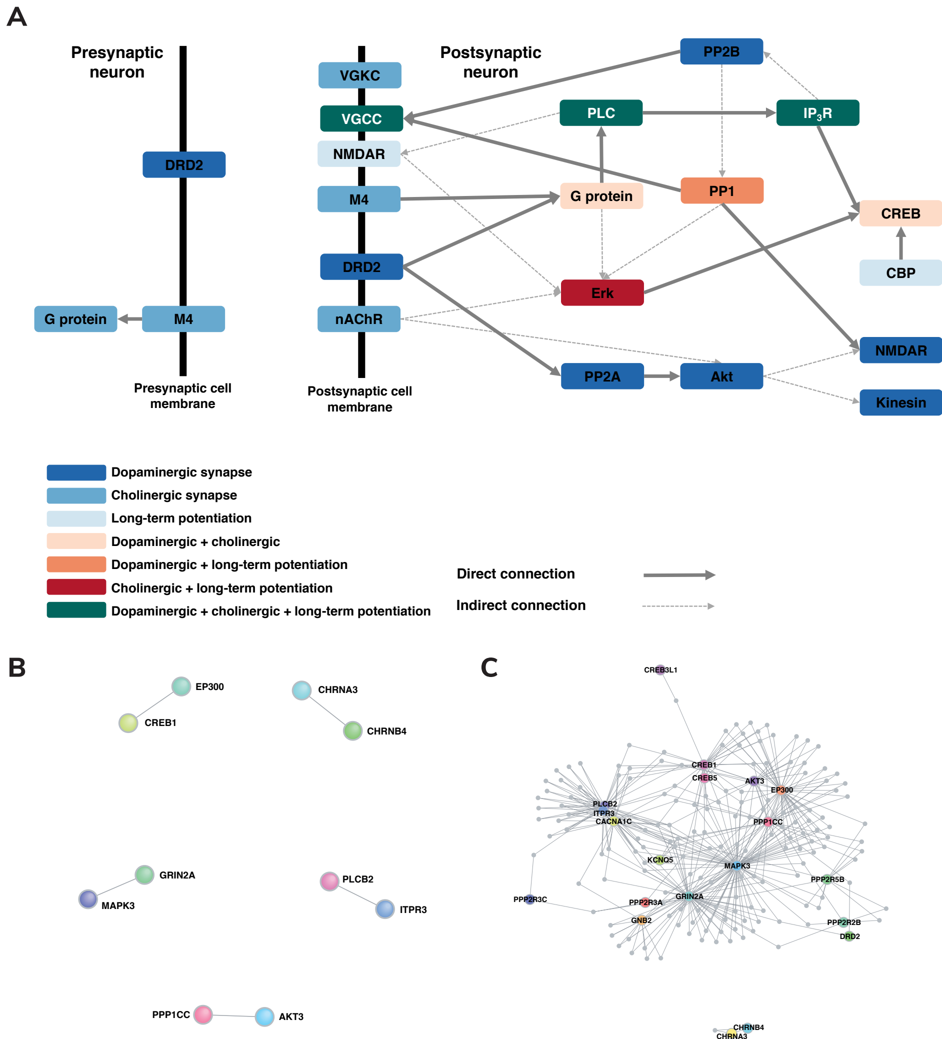


Figure 3. Enrichment and protein-protein interactions in KEGG synaptic signaling pathways. A) Significant pathway components (gene p-value < 1.45×10^{-4}) for dopaminergic synapse, cholinergic synapse and long-term potentiation. Colors indicate specificity to one or more pathways. Arrows indicate connections between signaling components (solid, direct interaction; dashed, indirect interaction), based on KEGG [20]. **B)** Direct (interaction p = 1.4×10^{-2}) and **C)** indirect (interaction p = 4.0×10^{-3}) protein-protein interaction networks generated with DAPPLE [11] for 22 top genes from the dopaminergic synapse, long-term potentiation and cholinergic synapse gene sets.

Abbreviations: DRD2, dopamine receptor D2; M4, muscarinic acetylcholine receptor M4; G protein, guanine nucleotide-binding protein; VGKC, voltage-gated potassium channel; VGCC, voltage-gated calcium channel; NMDAR, N-methyl-d-aspartate receptor; nAChR, nicotinic acetylcholine receptor; PP2A, protein phosphatase 2A; Akt, protein kinase B; Erk, extracellular signal-regulated kinase; CREB, cAMP response element binding protein; CBP, CREB-binding protein; PP1, protein-phosphatase 1; PLC, phospholipase C; IP_3R, inositol trisphosphate receptor; PP2B, protein-phosphatase 2B.

been reported as highly associated to psychiatric disease [9]. Possible reasons for this lack of replication are the larger sample size and the exclusion of other phenotypes than schizophrenia in the current study. We followed our first findings up in a targeted analysis on pathways representing synaptic signaling in all major neurotransmitter systems and demonstrated enrichment of schizophrenia SNPs in the dopaminergic synapse, long-term potentiation through the glutamatergic system and the cholinergic synapse, although the latter could not be confirmed using INRICH. We then showed that protein products of significant schizophrenia-associated genes in our enriched KEGG pathways show on average a higher rate of interactions with other proteins than expected by chance. Finally, we highlight that expression patterns of top genes from the enriched KEGG pathways do not substantially differ between brain regions and that enrichment levels are not substantially attenuated by including brain gene expression as a covariate in our MAGMA analysis.

Dysfunctional synaptic transmission impacts synaptic plasticity and brain development, mediated through long-term potentiation (LTP) and long-term depression (LTD) [24]. Although all five major neurotransmitter systems (dopamine, gamma-aminobutyric acid, glutamate, serotonin, and acetylcholine) have been implicated in schizophrenia, the extent to which each of them is involved had remained elusive [25,26]. Our results strongly support the involvement of the dopaminergic system, which has been extensively examined in schizophrenia. Previous studies have reported increased dopamine synthesis and release, and increased dopamine receptor expression in schizophrenia [25,27]. *DRD2* genetic variants are also implicated in schizophrenia and several of its intermediate phenotypes [28,29]. We here confirm an accumulation of *DRD2* genetic variants in schizophrenia and in signaling cascades downstream of this receptor. Enrichment of the glutamate-induced LTP pathway was another finding that could be verified using all enrichment analysis tools. Mediation of LTP is, however, not limited to the glutamatergic system as post-synaptic signaling molecules such as the above mentioned CREB, IP_3R and PKB mediate synaptic plasticity in other neurotransmitter systems (e.g. the dopaminergic system). Multiple lines of evidence link LTP to cognitive deficits in schizophrenia [30]. Cholinergic transmission may also be relevant to symptomatology of schizophrenia, especially in light of the high rates of nicotine abuse and a range of cognitive symptoms [31,32] which have led some to

postulate that schizophrenia is primarily a cognitive illness [33]. The implication of acetylcholine in schizophrenia is further supported by a landmark study investigating chromatin interactions between enhancer regions containing schizophrenia-associated loci and promoter regions of target genes [34]. In a broader perspective, our findings are in line with rare variant studies in schizophrenia that report mutations in genes coding for postsynaptic signaling components [35-38].

Our detailed analysis of downstream signaling cascades in all major neurotransmitter system gene sets revealed several of these cascades to be highly enriched for schizophrenia-associated variants: the phospholipase pathway, CREB signaling and the PKB/Akt signaling cascade. All of these cascades may be linked to schizophrenia by numerous lines of neurobiological evidence, as outlined below. First, the phospholipase pathway (particularly PLC) controls neuronal activity and thereby maintains synaptic functioning and development. Gene deletions in *PLC* are associated with schizophrenia and altered expression of *Plc* and schizophrenia-like behavior has been reported in *Plc* knock-out mice [39,40]. Second, signaling through CREB modulates synaptic plasticity. A recent study focusing on the cyclic adenosine monophosphate (cAMP)/PKA/CREB pathway shows a significant association of a SNP in this system with schizophrenia [41]. Additionally, ERK is part of the CREB signaling cascade and has been found to be enriched in our analyses. Impairment of signaling through ERK is hypothesized to constitute a disease mechanism in schizophrenia [42,43]. Third, we found a significant enrichment of schizophrenia SNPs in postsynaptic protein kinase B (PKB or Akt). *AKT1* messenger RNA levels are higher in blood of schizophrenia patients compared to healthy controls and interactions between genetic variation in *AKT1* and cannabis use are associated with schizophrenia, possibly mediated through AKT signaling downstream of DRD2 [44,45]. Interestingly, phosphorylation of glycogen synthase kinase 3 beta (Gsk3β) by the antipsychotic aripiprazole is mediated by Akt [46]. Finally, we detected an accumulation of SNPs in protein phosphatase 1 (*PP1*) and protein phosphatase 2A (*PP2A*). PP2A is one of the mediators of sensorimotor gating, an intermediate phenotype for schizophrenia [47].

Conditional analysis correcting for overlapping genes between gene sets and pathways indicated a high sharing of genes contributing to enrichment between

the two GO gene sets. That these GO gene sets are overlapping and were the only two being significantly enriched, implies a highly specific signal found through this analysis in MAGMA. Furthermore, these gene sets connect well to the analyzed KEGG pathways. Between KEGG pathways, conditioning on shared genes resulted in enrichment significance levels dropping to non-significance, supporting the hypothesis that signaling mechanisms shared between the implicated pathways are likely to be a strong underlying factor in schizophrenia in addition to neurotransmitter-specific pathways. Additionally, the *in silico* finding of higher interaction rates of proteins coded by SNP-enriched genes in the dopaminergic, LTP and cholinergic pathways hint at an important biological role of these genes in various processes, including synaptic signaling. However, this interaction may be expected since these proteins all operate in the same cellular compartment (the synapse). Expression profiles of top genes from enriched KEGG pathways among brain tissues did not show major differences, similarly to brain tissues versus other tissues (Supplementary Figure 7, Supplementary Figure 8). This could support the widespread role of these genes in biological processes. Seemingly higher relative transcript abundance of these genes in cerebellum and cerebellar hemisphere might be explained by the fact that these genes were selected from neuronal gene sets and that neuron density is higher in these tissues compared to the rest of the brain [48]. The relative lack of attenuation of enrichment in GO gene sets and KEGG pathways when conditioned on brain gene expression might not have been expected given that these pathways represent processes and systems in the brain. However, genes with low expression levels can still have a biologically relevant function. Furthermore, gene expression in the GTEx database has been quantified in miscellaneous samples, in whom gene expression levels could differ from a sample limited to schizophrenia subjects. Finally, these gene expression levels were derived from whole brain tissue containing different cell types, whereas our tested pathways only represented a specific cellular compartment.

Several limitations should be considered when interpreting our results. First, as our analyses are dependent on the power of GWASs, we cannot rule out the possibility that increased sample sizes in future studies may flag other synaptic systems also hypothesized to be associated with schizophrenia, such as the glutamatergic system [49], and as further illustrated by results pointing towards

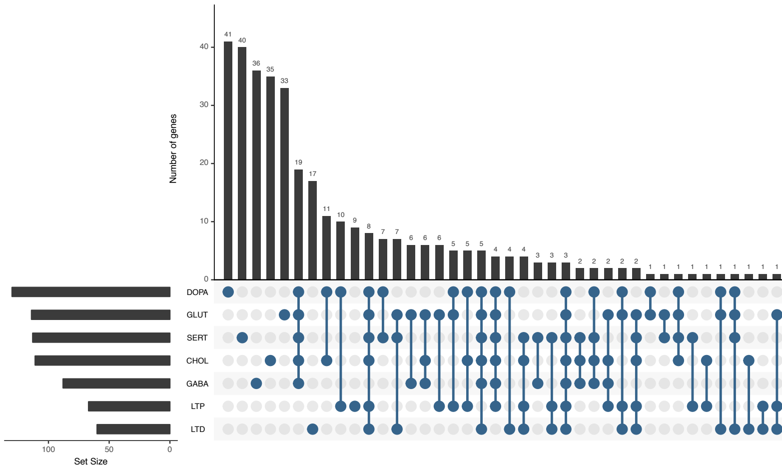
enrichment in the GABAergic synapse in novel GWAS [4] (Supplementary Figure 10). Second, we can only test for enrichment in gene sets and pathways that are annotated based on the knowledge currently available. Third, only protein-coding regions of the genome and up- and downstream regions in close proximity to genes were considered in our analyses. Non-coding stretches of the genome account for a major part of disease heritability and transcription regulation [34]. Advances in annotation of the non-coding parts of the genome will likely allow for future integration of these regions in currently available gene set enrichment analysis tools and novel functional genomics tools. These may be integrated with expression quantitative trait locus (eQTL) data and genomic interactions, which potentially prime novel molecular mechanisms involved in the pathways underlying schizophrenia. Finally, MAGMA, MAGENTA and INRICH slightly differ in their approach of testing a generally similar null hypothesis - that genes in a given gene set are not more strongly associated with schizophrenia than genes outside such a gene set. Despite their disparaging analytical approaches but perhaps owing to these similar null hypotheses, we found highly similar results using these tools as would be expected.

In conclusion, using complementary enrichment analysis approaches, we highlight downstream signaling cascades as the most likely part of the dopaminergic, cholinergic and LTP systems to have a role in schizophrenia. Our results open avenues for further research aimed at elucidating signaling pathways in schizophrenia, e.g. through more comprehensive integration with genotype, expression and protein-interaction data. Finally, our findings may aid the discovery of novel drug targets by prioritizing neurotransmitter systems and their signaling molecules, e.g. through translational experiments or drug-interaction studies, to hopefully reduce the burden imposed on quality of life in patients suffering from this disabling disorder.

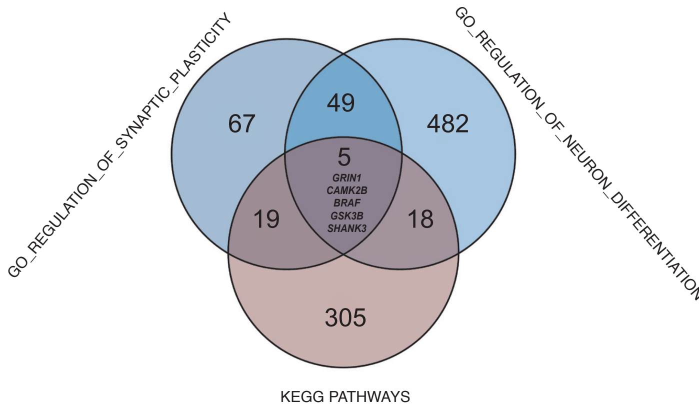
2.5 Supplementary information

2.5.1 *Supplementary figures*

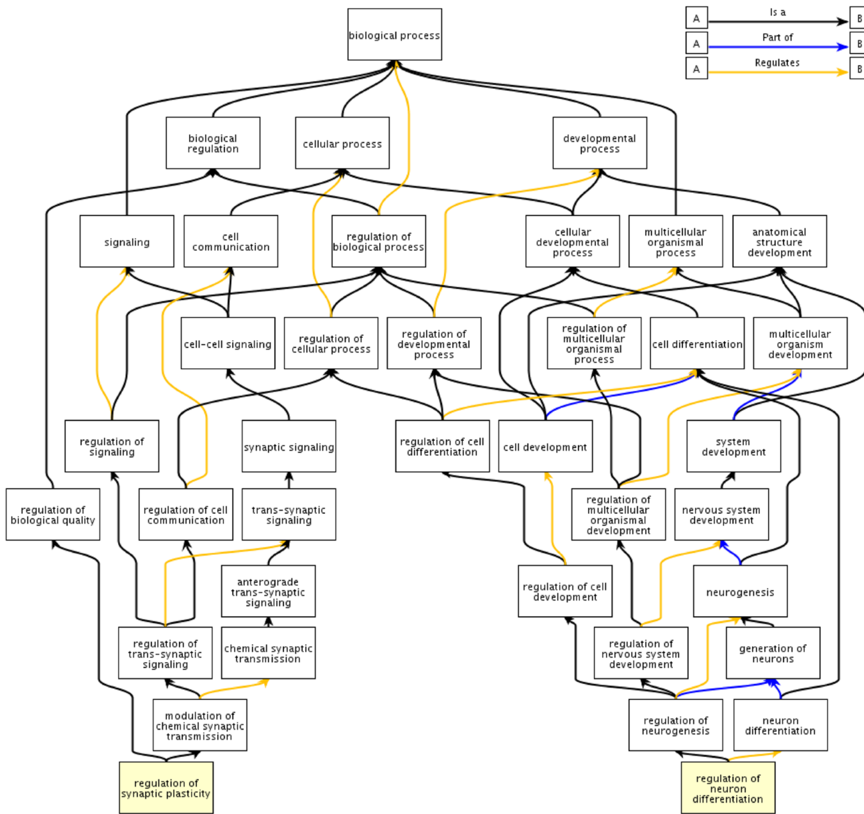
A



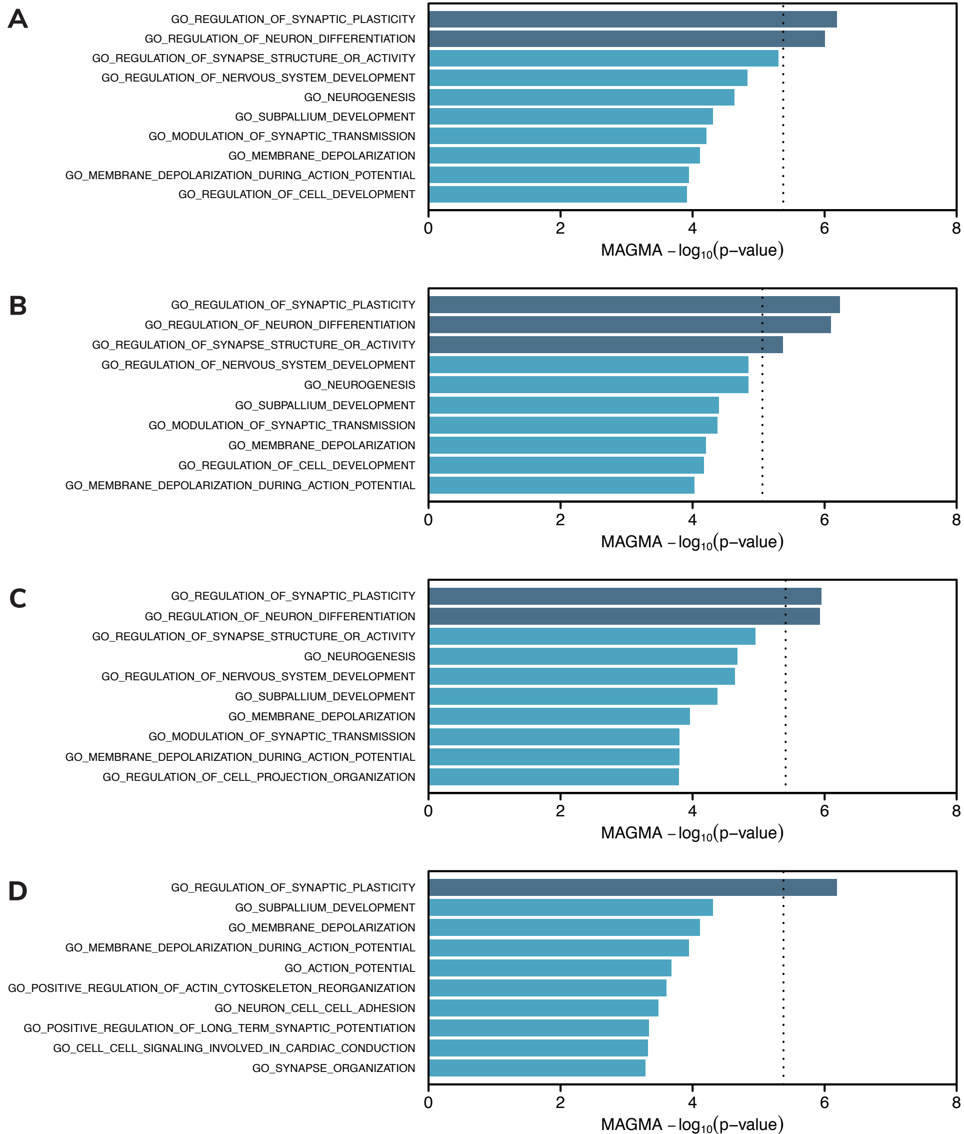
B



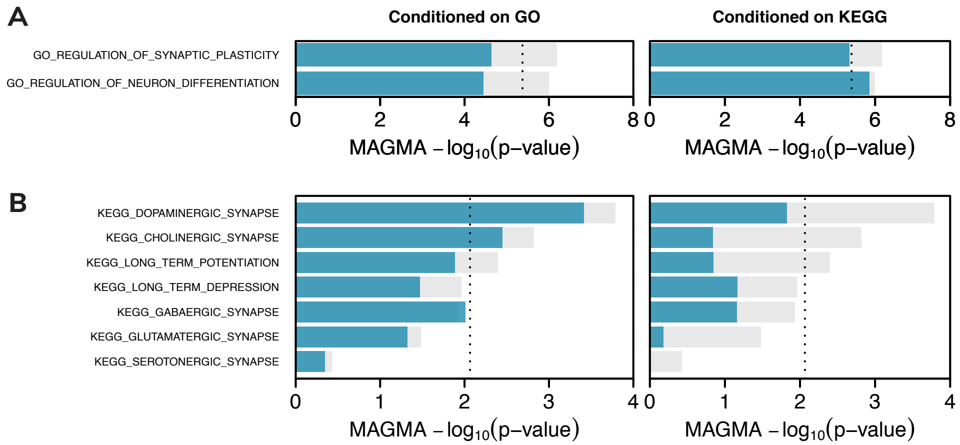
Supplementary Figure 1. Overlap between tested gene sets. A) Overlap between KEGG synaptic signaling pathways. Gene set sizes are shown in the left horizontal bar plot. The top vertical bar plot (upper panel) shows the number of genes that are either unique to a pathway (indicated by single highlighted blue dot in the lower panel) or shared between multiple pathways (corresponding to connected highlighted blue dots in the lower panel). More than one third of the genes are shared between one or more pathways (136, vs. 211 that are unique to a pathway). **B)** Venn-diagram showing overlap between significant MSigDB GO gene sets and genes in tested KEGG synaptic signaling pathways. Five genes that overlap between all sets are shown in the middle segment. Abbreviations: DOPA, KEGG_DOPAMINERGIC_SYNAPSE; GLUT, KEGG_GLUTAMATERGIC_SYNAPSE; SERT, KEGG_SEROTONERGIC_SYNAPSE; CHOL, KEGG_CHOLINETGIC SYNAPSE; GABA, KEGG_GABAERGIC_SYNAPSE; LTP, KEGG_LONG_TERM_POTENTIATION; LTD, KEGG_LONG_TERM_DEPRESSION. Graph in (A) made using *UpSetR* package in R version 3.3.3 [50]. Graph in (B) made using *VennDiagram* package in R version 3.3.3 [51].



Supplementary Figure 2. Gene Ontology ancestor chart of significantly enriched biological processes. GO biological processes significantly enriched with schizophrenia SNPs are marked in yellow ('Regulation of neuron differentiation' and 'Regulation of synaptic plasticity'). These are specialized terms of parent terms describing synaptic signaling and neural development. Figure generated using *QuickGO* [52]. Arrow colors are explained in the legend.

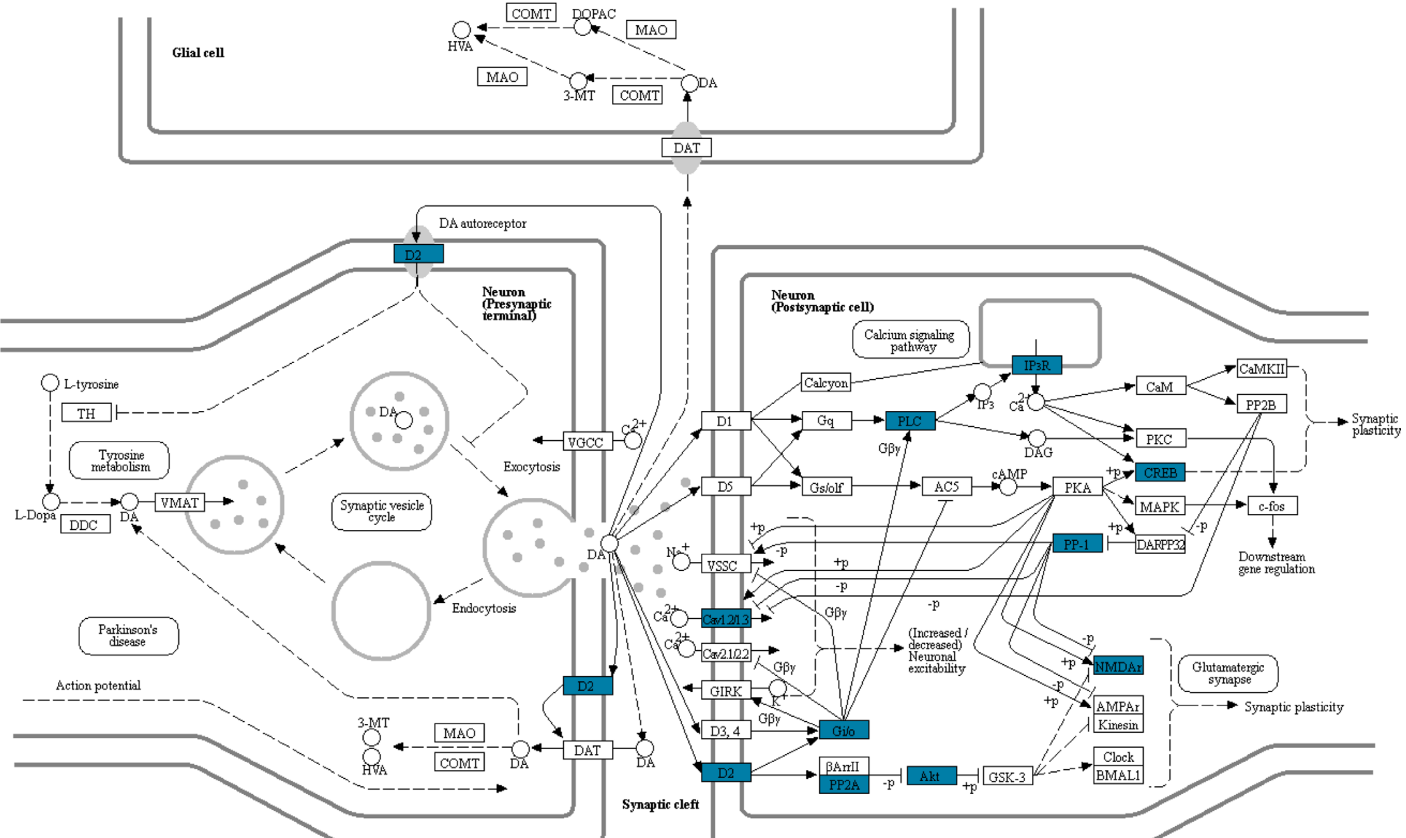


Supplementary Figure 3. Sensitivity analysis results for MSigDB GO biological processes. Enrichment $-\log_{10}$-converted p-values are shown for GO gene sets in **A)** the main analysis corresponding to Figure 2A in the main text, and compared to this analysis the results for analyses with **B)** the extended MHC region (chr6:25,500,000–33,500,000) excluded (significance threshold $p = 8.75 \times 10^{-6}$, dotted line), **C)** excluding the X chromosome from the analysis (significance threshold $p = 3.90 \times 10^{-6}$, dotted line), and **D)** strict size filtering of GO gene sets ($10 < n_{genes} < 200$) (significance threshold $p = 4.20 \times 10^{-6}$, dotted line).

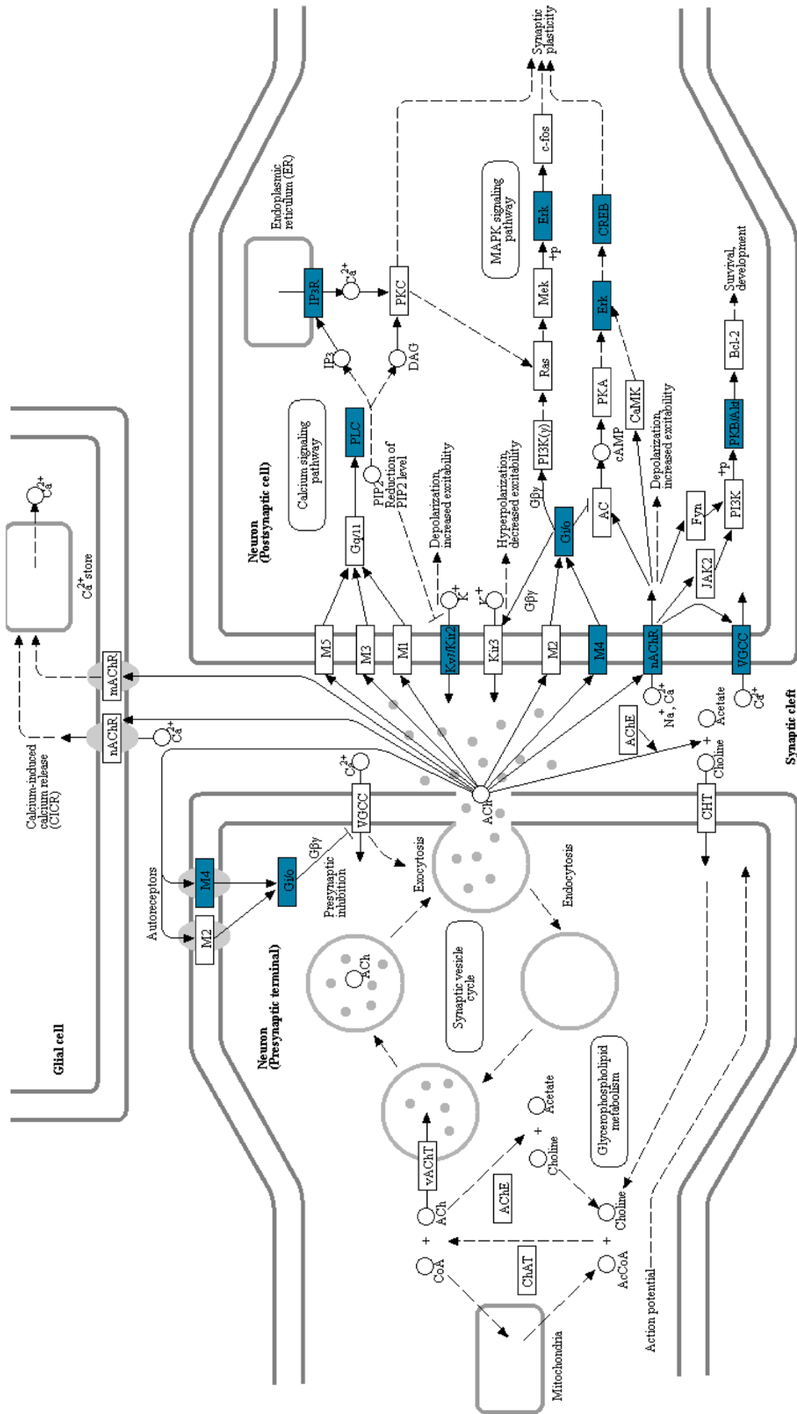


Supplementary Figure 4. Conditional gene set enrichment analysis in significant MSigDB GO gene sets and KEGG synapse pathways. A) Enrichment in significant GO gene sets was conditioned on shared genes between the significant GO gene sets (blue bars, left panel) and shared genes with the tested KEGG pathways (blue bars, right panel). $-\log_{10}$-converted enrichment p-values of the initial analysis are shown in gray (corresponding to Figure 2A in the main text), dotted lines indicate multiple-testing corrected significance threshold ($p < 4.23 \times 10^{-6}$). **B)** Enrichment in KEGG pathways was conditioned on shared genes with significant GO gene sets (blue bars, left panel) and shared genes with other tested KEGG pathways (blue bars, right panel). $-\log_{10}$-converted enrichment p-values of the initial analysis are shown in gray (corresponding to Figure 2B in the main text), dotted lines indicate the multiple-testing corrected significance threshold ($p < 8.6 \times 10^{-3}$).

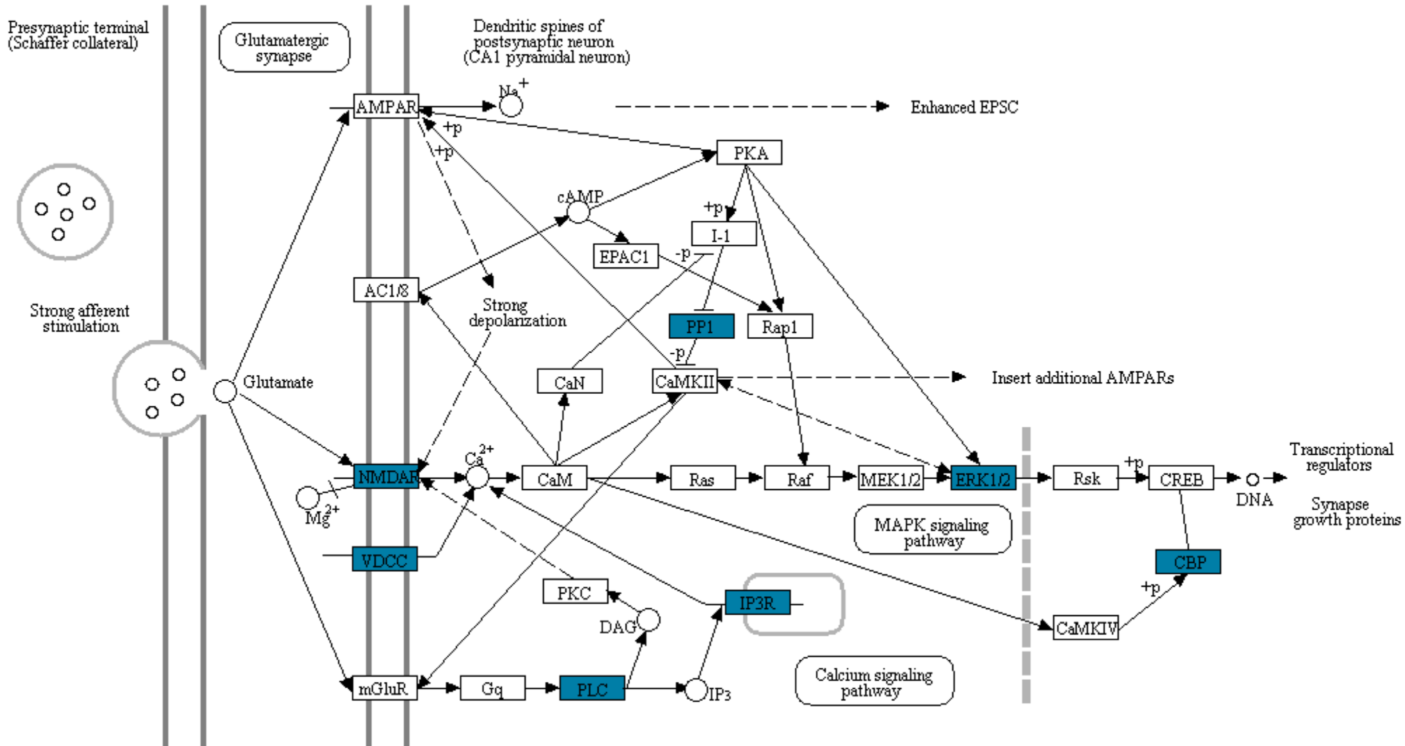
Supplementary Figure 5A



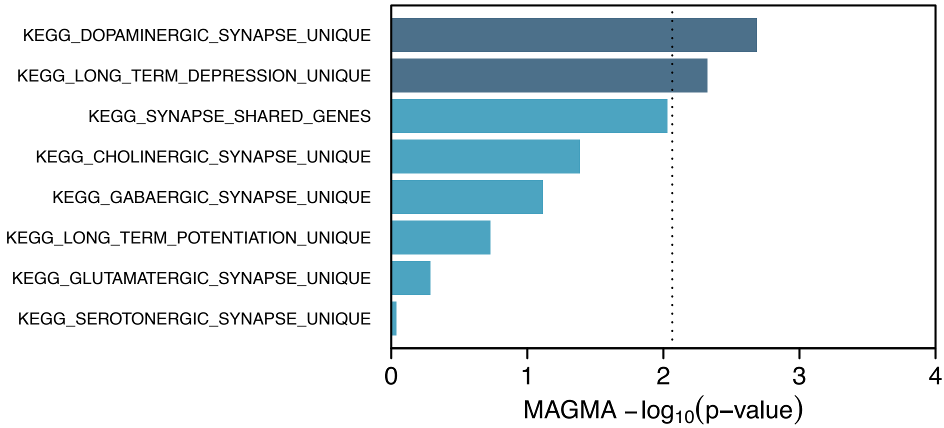
Supplementary Figure 5B



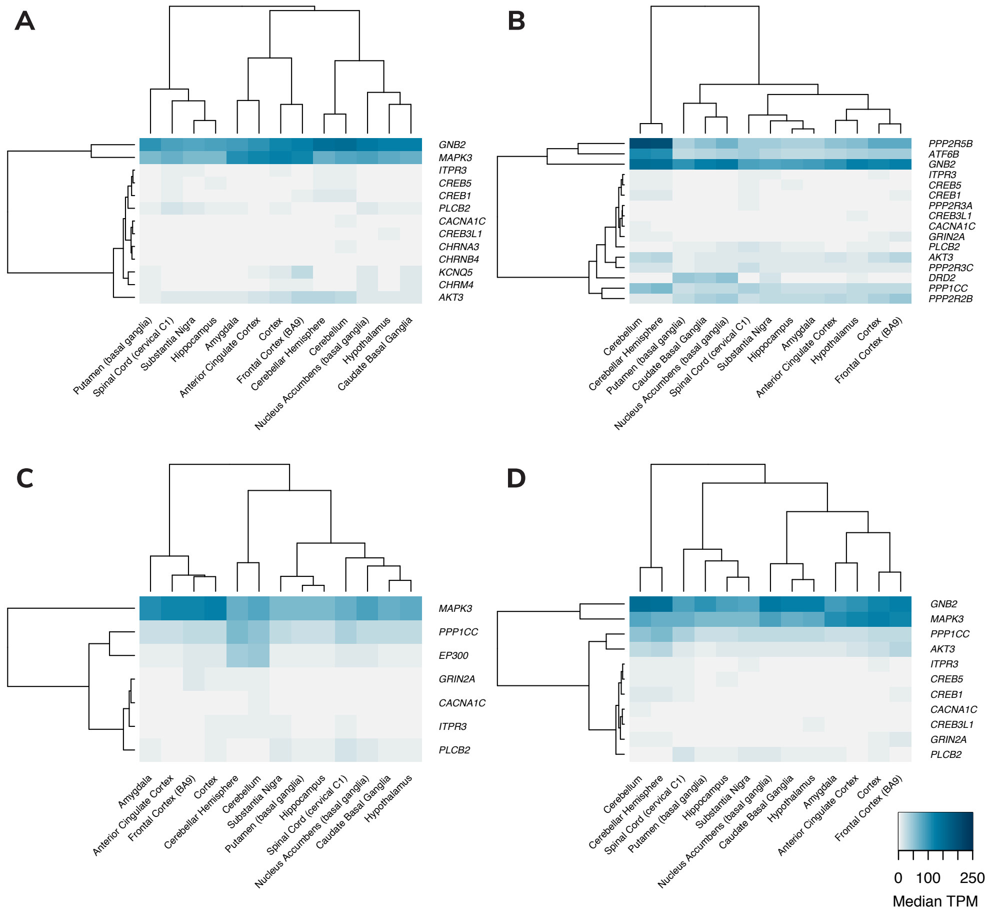
Supplementary Figure 5C



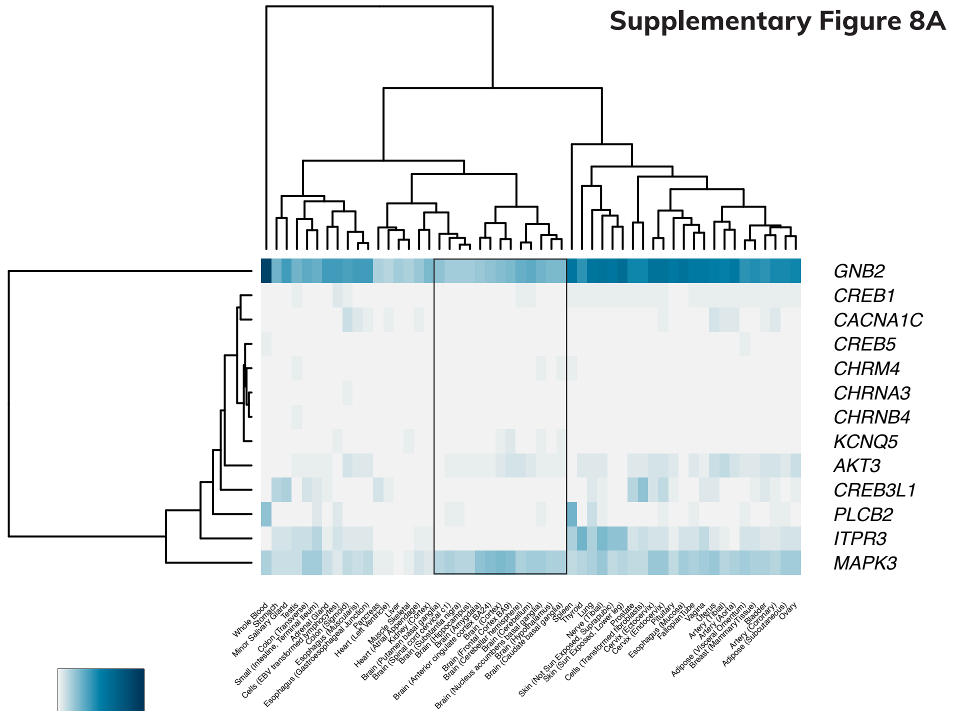
Supplementary Figure 5. Localization of enriched genes in KEGG synaptic signaling pathways. KEGG pathway components containing significant schizophrenia-associated genes, selected from 344 unique genes included in the tested KEGG pathways (gene $p < 0.05/344 = 1.45 \times 10^{-4}$), are highlighted in blue for the three significantly enriched pathways: **A)** Dopaminergic synapse **B)** Cholinergic synapse and **C)** Long-term potentiation. These pathways were derived from KEGG [20] and highlighted using the *pathview* package in R [21].

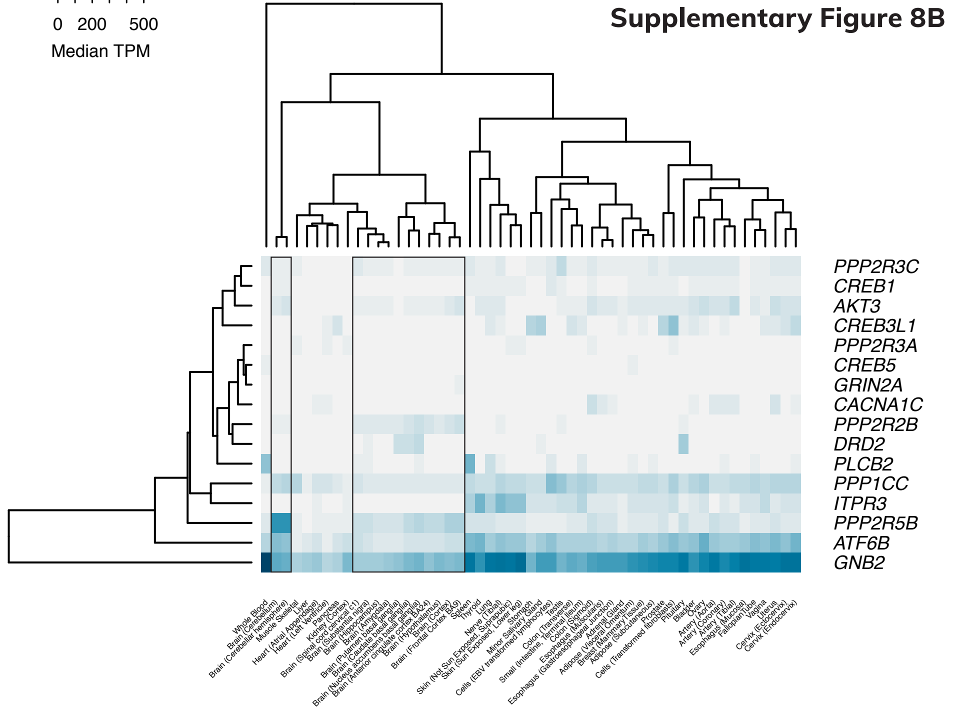


Supplementary Figure 6. Enrichment in KEGG pathways stratified on unique and shared genes. Enrichment $-\log_{10}$-converted p-values are shown for KEGG pathways stratified on unique (*_UNIQUE) and overlapping (KEGG_SYNAPSE_SHARED_GENES) genes between the analyzed gene sets. Significance threshold (indicated by the dotted line) was similar to the main analysis ($p < 8.6 \times 10^{-3}$).



Supplementary Figure 7. Median relative transcript abundance of top genes in enriched KEGG pathways in 13 GTEx brain tissues. Heat maps show median transcript abundance of top genes in enriched KEGG pathways: **A)** Cholinergic synapse, **B)** Dopaminergic synapse, **C)** Long-term potentiation, **D)** Shared genes (genes present in two or three of the significantly enriched KEGG gene sets). Colors indicate median Transcript Per Million (TPM) values (gray is low, blue is high, TPM range from 0-250 in each heatmap). Genes and tissues are clustered based on similarities in relative transcript abundance. Gene TPM data was obtained from GTEx [12], heat maps were generated with R-package *gplots* [23].


Supplementary Figure 8A
GNB2
CREB1
CACNA1C
CREB5
CHRM4
CHRNA3
CHRNB4
KCNQ5
AKT3
CREB3L1
PLCB2
ITPR3
MAPK3


0 200 500
Median TPM
Supplementary Figure 8B
PPP2R3C
CREB1
AKT3
CREB3L1
PPP2R3A
CREB5
GRIN2A
CACNA1C
PPP2R2B
DRD2
PLCB2
PPP1CC
ITPR3
PPP2R5B
ATF6B
GNB2

Supplementary Figure 8C

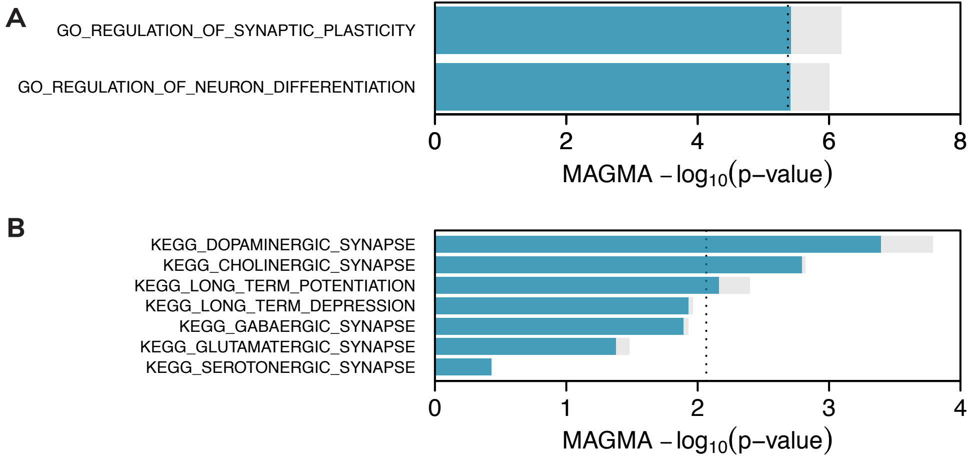
PPP1CC
EP300
ITPR3
PLCB2
CACNA1C
GRIN2A
MAPK3

0 200 500
Median TPM

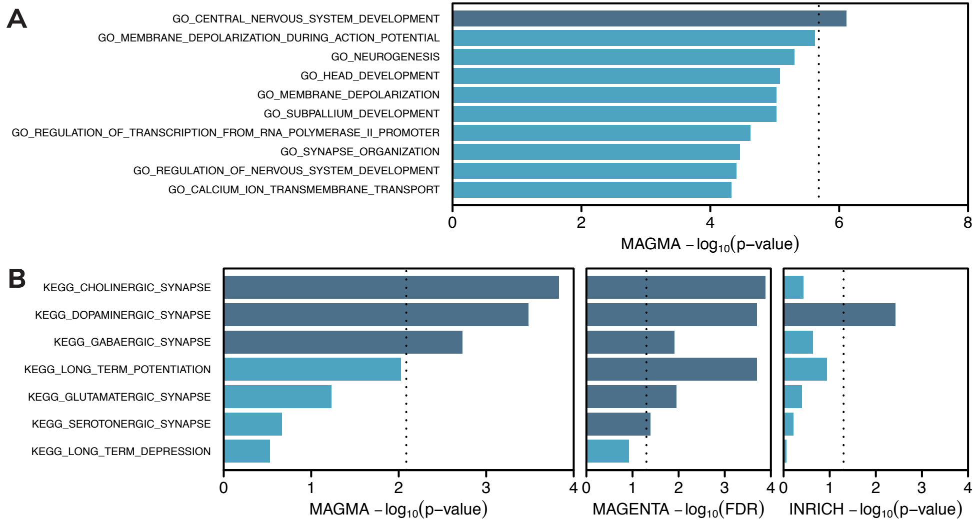
Supplementary Figure 8D

PPP1CC
ITPR3
CREB1
CACNA1C
CREB5
GRIN2A
AKT3
CREB3L1
PLCB2
MAPK3
GNB2

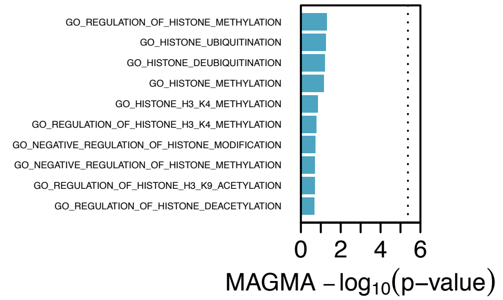
< Previous two pages - **Supplementary Figure 8. Median relative transcript abundance of top genes in enriched KEGG pathways in 53 GTEx tissues.** Heat maps show median transcript abundance of top genes in enriched KEGG pathways: **A)** Cholinergic synapse, **B)** Dopaminergic synapse, **C)** Long-term potentiation, **D)** Shared genes (genes present in two or three of the enriched KEGG gene sets). Colors indicate median Transcript Per Million (TPM) values (gray is low, blue is high, TPM range from 0-500 in each heat map). Genes and tissues are clustered based on similarities in relative transcript abundance. Brain tissue clusters are marked with black boxes. Gene TPM data was obtained from GTEx [12], heat maps generated with R-package *gplots* [23].

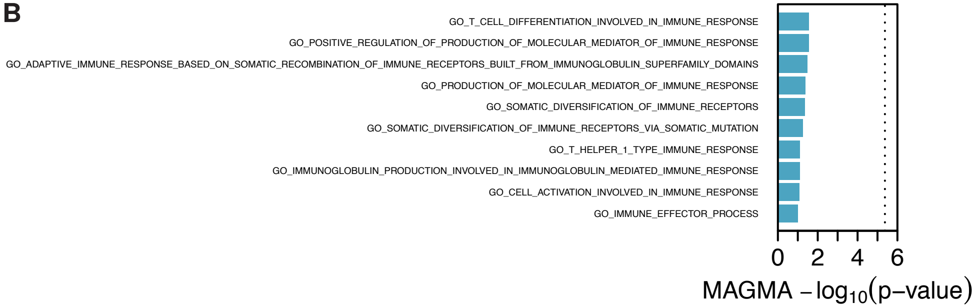


Supplementary Figure 9. Conditional analysis with brain gene expression levels as covariate. A) Enrichment in significant GO gene sets was conditioned on brain gene expression (blue bars). -$\log_{10}$-converted enrichment p-values of the initial analysis are shown in gray (corresponding to Figure 2A in the main text), dotted line indicates the multiple-testing corrected significance threshold ($p < 4.23 \times 10^{-6}$). **B)** Enrichment in KEGG pathways was conditioned on brain gene expression (blue bars). -$\log_{10}$-converted enrichment p-values of the initial analysis are shown in gray (corresponding to Figure 2B in the main text), dotted line indicates the multiple-testing corrected significance threshold ($p < 8.6 \times 10^{-3}$).



Supplementary Figure 10. Results of enrichment analysis on MSigDB biological processes and KEGG pathways for the PGC 2014 + CLOZUK meta-analysis [4]. Reported p-values or FDR are $-\log_{10}$-converted. Significance thresholds are different per analysis depending on the used method. Significant enrichment is indicated with dark blue. **A)** Top ten enriched MSigDB GO gene sets (significance threshold $p = 2.07 \times 10^{-6}$, dotted line). **B)** Enrichment in KEGG synaptic signaling pathways tested using MAGMA (left panel, significance threshold $p = 8.2 \times 10^{-3}$, dotted line), MAGENTA (middle panel, significance threshold FDR = 0.05, dotted line) and INRICH (right panel, significance threshold p = 0.05, dotted line).





Supplementary Figure 11. Results for histone and immune related GO gene sets. A) Histone and **B)** immune related gene sets were found to be associated with psychiatric disorders in an earlier publication [9]. These gene sets did not reach significance in our primary enrichment analysis of GO gene sets. Results for the ten most enriched histone and immune pathways are shown as $-\log_{10}$-converted enrichment p-values derived from the main analysis result (Figure 2A in the main text and Supplementary Table 1). Significance threshold is similar to the main analysis ($p < 4.23 \times 10^{-6}$).

2.5.2 *Supplementary tables*

Supplementary Table 1. Top 50 enriched MSigDB Gene Ontology gene sets in MAGMA analysis.

Gene Set	N Genes	Beta	Std Beta	SE	P
GO_REGULATION_OF_SYNAPTIC_PLASTICITY	140	0.45	0.0385	0.093	6.52E-07
GO_REGULATION_OF_NEURON_DIFFERENTIATION	551	0.243	0.0407	0.051	1.00E-06
GO_REGULATION_OF_SYNAPSE_STRUCTURE_OR_ACTIVITY	231	0.329	0.036	0.0745	5.03E-06
GO_REGULATION_OF_NERVOUS_SYSTEM_DEVELOPMENT	745	0.181	0.0351	0.0434	1.48E-05
GO_NEUROGENESIS	1393	0.131	0.0341	0.0322	2.34E-05
GO_SUBPALLIUM_DEVELOPMENT	22	1.01	0.0343	0.259	4.94E-05
GO_MODULATION_OF_SYNAPTIC_TRANSMISSION	297	0.251	0.031	0.0653	6.21E-05
GO_MEMBRANE_DEPOLARIZATION	61	0.548	0.031	0.145	7.82E-05
GO_MEMBRANE_DEPOLARIZATION_DURING_ACTION_POTENTIAL	39	0.65	0.0294	0.176	0.00011386
GO_REGULATION_OF_CELL_DEVELOPMENT	832	0.15	0.0306	0.0408	0.00012223
GO_POSITIVE_REGULATION_OF_NEURON_DIFFERENTIATION	304	0.248	0.0311	0.0683	0.00013929
GO_REGULATION_OF_CELL_PROJECTION_ORGANIZATION	553	0.18	0.0303	0.0502	0.00016298
GO_REGULATION_OF_TRANSMEMBRANE_TRANSPORT	422	0.199	0.0292	0.0556	0.000179
GO_ACTION_POTENTIAL	94	0.416	0.0291	0.118	0.00020884
GO_POSITIVE_REGULATION_OF_NERVOUS_SYSTEM_DEVELOPMENT	433	0.2	0.0298	0.0573	0.00023857
GO_POSITIVE_REGULATION_OF_ACTIN_CYTOSKELETON_REORGANIZATION	16	0.96	0.0278	0.276	0.00024819
GO_REGULATION_OF_NEURON_PROJECTION_DEVELOPMENT	405	0.207	0.0299	0.0599	0.00026923
GO_NEURON_CELL_CELL_ADHESION	15	1.1	0.0308	0.323	0.00032987
GO_POSITIVE_REGULATION_OF_LONG_TERM_SYNAPTIC_POTENTIATION	14	0.976	0.0264	0.295	0.00046483
GO_CELL_CELL_SIGNALING_INVOLVED_IN_CARDIAC_CONDUCTION	22	0.809	0.0275	0.245	0.00047746
GO_SYNAPSE_ORGANIZATION	143	0.339	0.0293	0.103	0.00051685
GO_REGULATION_OF_TRANSPORT	1782	0.0908	0.0264	0.0281	0.00061924
GO_MEMBRANE_ORGANIZATION	870	0.127	0.0264	0.0392	0.00062095
GO_REGULATION_OF_TRANSPORTER_ACTIVITY	198	0.256	0.026	0.0794	0.0006308
GO_POSITIVE_REGULATION_OF_SYNAPSE_MATURATION	11	1.25	0.0299	0.399	0.00090013
GO_REGULATION_OF_CALCIUM_ION_TRANSMEMBRANE_TRANSPORT	115	0.323	0.025	0.105	0.0010133
GO_REGULATION_OF_CELLULAR_LOCALIZATION	1261	0.1	0.0249	0.0325	0.0010238
GO_POSITIVE_REGULATION_OF_CELL_DEVELOPMENT	469	0.163	0.0253	0.0545	0.0013484
GO_REGULATION_OF_SKELETAL_MUSCLE_CELL_DIFFERENTIATION	17	0.826	0.0247	0.276	0.0013846
GO_NEGATIVE_REGULATION_OF_CATION_CHANNEL_ACTIVITY	34	0.548	0.0231	0.184	0.0014228

< Supplementary Table 1 continued.

GO_ESTABLISHMENT_OF_LOCALIZATION_IN_CELL	1646	0.0875	0.0246	0.0293	0.0014271
GO_CENTRAL_NERVOUS_SYSTEM_DEVELOPMENT	866	0.117	0.0244	0.0395	0.0014967
GO_NEGATIVE_REGULATION_OF_NITROGEN_COMPOUND_METABOLIC_PROCESS	1485	0.0922	0.0247	0.0311	0.0015245
GO_MEMBRANE_DEPOLARIZATION_DURING_CARDIAC_MUSCLE_CELL_ACTION_POTENTIAL	14	0.912	0.0247	0.308	0.0015471
GO_HIPPOCAMPUS_DEVELOPMENT	73	0.427	0.0264	0.145	0.0016643
GO_REGULATION_OF_CATION_CHANNEL_ACTIVITY	88	0.352	0.0238	0.12	0.0016716
GO_SYNAPSE_ASSEMBLY	67	0.475	0.0281	0.165	0.0020085
GO_NEUROMUSCULAR_SYNAPTIC_TRANSMISSION	26	0.661	0.0244	0.23	0.0020214
GO_POSITIVE_REGULATION_OF_REGULATED_SECRETORY_PATHWAY	48	0.465	0.0233	0.163	0.0021383
GO_NEURON_MIGRATION	110	0.314	0.0238	0.11	0.0022175
GO_NEGATIVE_REGULATION_OF_TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER	736	0.122	0.0235	0.043	0.0022974
GO_LONG_TERM_SYNAPTIC_POTENTIATION	39	0.511	0.0231	0.181	0.0023777
GO_CALCIUM_ION_TRANSMEMBRANE_TRANSPORT	158	0.258	0.0234	0.0921	0.0025637
GO_METENCEPHALON_DEVELOPMENT	100	0.306	0.0221	0.11	0.0026063
GO_SYNAPTIC_SIGNALING	420	0.154	0.0226	0.055	0.0026091
GO_POSITIVE_REGULATION_OF_NEURON_PROJECTION_DEVELOPMENT	230	0.222	0.0242	0.0795	0.0026199
GO_CELL_FATE_SPECIFICATION	70	0.384	0.0232	0.138	0.0026895
GO_REGULATION_OF_ORGANELLE_ORGANIZATION	1160	0.096	0.0229	0.0345	0.0027027
GO_FOREBRAIN_NEURON_FATE_COMMITMENT	12	0.856	0.0215	0.311	0.0029888
GO_REGULATION_OF_NEURONAL_SYNAPTIC_PLASTICITY	49	0.427	0.0216	0.156	0.0030554

Multiple testing-corrected significance threshold: $p = 4.2 \times 10^{-6}$

Gene set: Name of the Gene Ontology gene set from MSigDB. N Genes: Number of genes in the Gene Ontology gene set that were used in the analysis. Beta: Regression coefficient of the Gene Ontology gene set. Std Beta: The semi-standardized regression coefficient, corresponding to the predicted change in Z-value given a change of one standard deviation in the predictor gene set. SE: Standard error of the regression coefficient. P: Competitive gene-set p-value.

This table shows the results of Figure 2A. We performed an enrichment analysis with schizophrenia-associated variants in GO gene sets derived from MSigDB, a database where GO terms are collected and combined into single gene sets when they show large overlap. In total, 4436 GO gene sets were tested using a competitive gene set analysis in MAGMA.

Supplementary Table 2. Result MAGMA analysis in KEGG synaptic signaling gene sets.

KEGG Pathway	N Genes in set	N Genes in analysis	Beta	Std Beta	SE	P
KEGG_DOPAMINERGIC_SYNAPSE	130	128	0.367	0.03	0.102	0.00016191
KEGG_CHOLINERGIC_SYNAPSE	111	111	0.323	0.0246	0.109	0.0015128
KEGG_LONG_TERM_POTENTIATION	67	67	0.344	0.0204	0.13	0.0040026
KEGG_LONG_TERM_DEPRESSION	60	60	0.321	0.018	0.14	0.010868
KEGG_GABAERGIC_SYNAPSE	88	88	0.288	0.0195	0.127	0.011796
KEGG_GLUTAMATERGIC_SYNAPSE	114	114	0.198	0.0153	0.108	0.033007
KEGG_SEROTONERGIC_SYNAPSE	113	112	0.0353	0.0027	0.108	0.37243

Multiple testing-corrected significance threshold: $p = 8.6 \times 10^{-3}$
KEGG pathway: Name of the KEGG pathway. N Genes in set: Total number of genes in the KEGG pathway. N Genes in analysis: The number of genes from KEGG pathway that was effectively used in the analysis. Beta: Regression coefficient of the KEGG pathway. Std Beta: The semi-standardized regression coefficient, corresponding to the predicted change in Z-value given a change of one standard deviation in the predictor gene set. SE: Standard error of the regression coefficient. P: Competitive gene-set p-value.

This table shows the results of Figure 2B. Using MAGMA, we tested for enrichment of schizophrenia-associated variants in seven gene sets derived from KEGG representing synaptic signaling processes in different neurotransmitter systems.

Supplementary Table 3. Top associated genes in tested KEGG pathways.

Gene ID (Entrez)	Gene symbol	Chr	BP Start	BP End	N SNPs	N Param	N Sets	CH	GL	DO	GA	SE	LP	LD	Z	P
775	*CACNA1C*	12	2059952	2827115	2156	183	6	1	1	1	1	1	1	0	8.4436	1.54E-17
1132	*CHRM4*	11	46385335	46433122	39	10	1	1	0	0	0	0	0	0	6.7241	8.84E-12
1143	*CHRNB4*	15	78896636	78953587	151	22	1	1	0	0	0	0	0	0	6.5813	2.33E-11
10000	*AKT3*	1	243631535	244034381	562	23	2	1	0	1	0	0	0	0	6.2318	2.31E-10
1813	*DRD2*	11	113260317	113366413	276	30	1	0	0	1	0	0	0	0	6.0238	8.52E-10
1136	*CHRNA3*	15	78865394	78933637	173	16	1	1	0	0	0	0	0	0	5.9546	1.30E-09
140679	*SLC32A1*	20	37333105	37378015	181	24	1	0	0	0	1	0	0	0	5.9213	1.60E-09
1388	*ATF6B*	6	32063045	32116017	64	23	1	0	0	1	0	0	0	0	5.9193	1.62E-09
5334	*PLCL1*	2	198649426	199034608	693	36	1	0	0	0	1	0	0	0	5.7869	3.58E-09
5330	*PLCB2*	15	40556461	40620174	169	41	6	1	1	1	0	1	1	1	5.7429	4.65E-09
3710	*ITPR3*	6	33567951	33684351	551	45	6	1	1	1	0	1	1	1	5.7398	4.74E-09
4842	*NOS1*	12	117625921	117819607	518	44	1	0	0	0	0	0	0	1	5.6257	9.24E-09
1565	*CYP2D6*	22	42502501	42546883	170	14	1	0	0	0	0	1	0	0	5.5142	1.75E-08
5595	*MAPK3*	16	30105426	30154630	54	13	5	1	1	0	0	1	1	1	5.1557	1.26E-07
5526	*PPP2R5B*	11	64672143	64721950	90	10	1	0	0	1	0	0	0	0	5.0808	1.88E-07
9568	*GABBR2*	9	101030364	101491479	1817	155	1	0	0	0	1	0	0	0	4.9082	4.59E-07
2033	*EP300*	22	41468614	41596081	205	20	1	0	0	0	0	0	1	0	4.7979	8.02E-07
5523	*PPP2R3A*	3	135664515	135886752	371	24	1	0	0	1	0	0	0	0	4.6975	1.32E-06
1385	*CREB1*	2	208374616	208490284	173	17	2	1	0	1	0	0	0	0	4.6717	1.49E-06
5521	*PPP2R2B*	5	145949067	146481083	1193	77	1	0	0	1	0	0	0	0	4.6662	1.53E-06
90993	*CREB3L1*	11	46279189	46362972	116	20	2	1	0	1	0	0	0	0	4.6393	1.75E-06

< Supplementary Table 3 continued.

Gene ID (Entrez)	Gene symbol	Chr	BP Start	BP End	N SNPs	N Param	N Sets	CH	GL	DO	GA	SE	LP	LD	Z	P
2913	*GRM3*	7	86253230	86514193	529	42	1	0	1	0	0	0	0	0	4.547	2.72E-06
1742	*DLG4*	17	7073209	7143369	106	24	1	0	1	0	0	0	0	0	4.4138	5.08E-06
2903	*GRIN2A*	16	9827265	10296611	2001	63	3	0	1	1	0	0	1	0	4.1802	1.46E-05
5592	*PRKG1*	10	52730911	54078110	4629	245	1	0	0	0	0	0	0	1	4.142	1.72E-05
10672	*GNA13*	17	62985407	63072920	134	11	1	0	0	0	0	0	0	1	4.0616	2.44E-05
5501	*PPP1CC*	12	111137613	111200783	145	15	2	0	0	1	0	0	1	0	3.9895	3.31E-05
2566	*GABRG2*	5	161474648	161602545	255	35	1	0	0	0	1	0	0	0	3.9564	3.80E-05
56479	*KCNQ5*	6	73311571	73928574	1681	114	1	1	0	0	0	0	0	0	3.8291	6.43E-05
2783	*GNB2*	7	100251363	100296792	77	14	5	1	1	1	1	1	0	0	3.7544	8.69E-05
2555	*GABRA2*	4	46226470	46412056	428	28	1	0	0	0	1	0	0	0	3.7351	9.38E-05
11337	*GABARAP*	17	7123738	7165753	53	9	1	0	0	0	1	0	0	0	3.716	1.01E-04
55012	*PPP2R3C*	14	35534678	35611868	264	21	1	0	0	1	0	0	0	0	3.6898	1.12E-04
9586	*CREB5*	7	28318940	28885511	1792	172	2	1	0	1	0	0	0	0	3.6343	1.39E-04

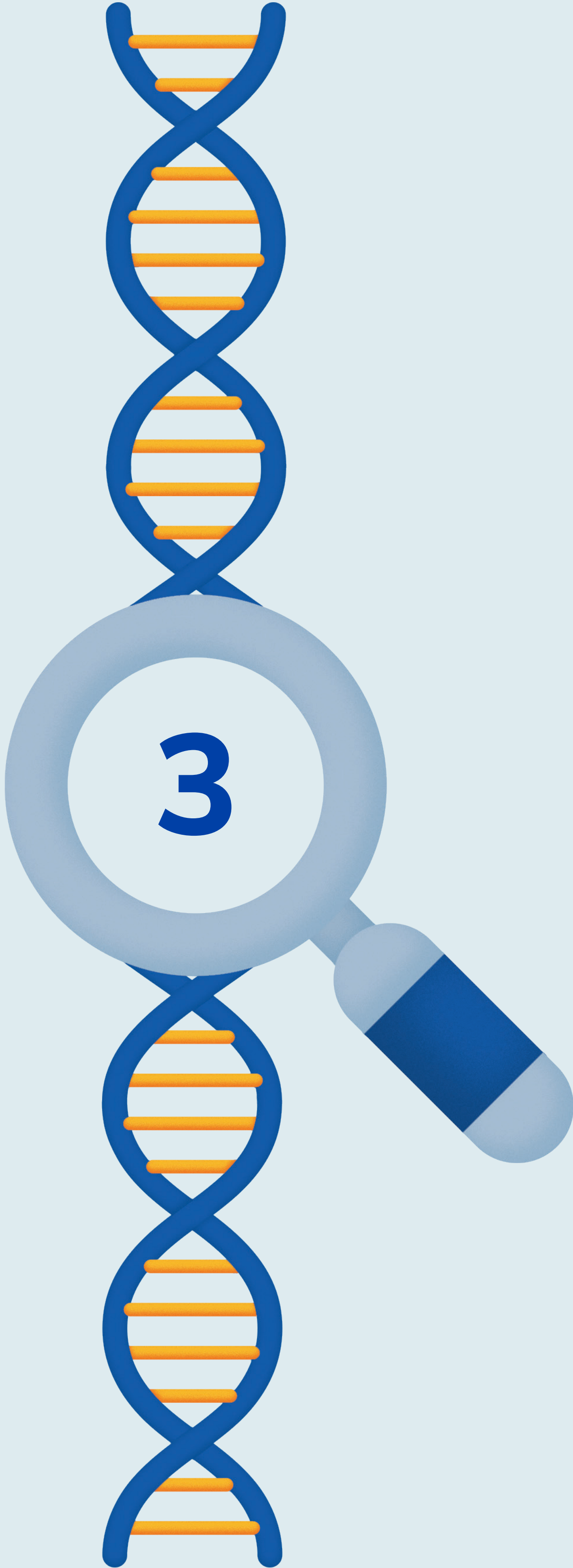
Genes with a gene p-value (from MAGMA gene analysis step) of $p < 0.05/344$ unique genes in tested KEGG gene sets ($p < 1.45 \times 10^{-4}$) are shown in this table.
Gene ID (Entrez) / Gene symbol: Gene Entrez identifier and gene symbol. Chr: Chromosome number. BP start: Basepair start position of the gene (hg19). BP end: Basepair end position of the gene (hg19). N SNPs: Number of SNPs mapped to the gene.
N Param: Number of relevant parameters used in the gene analysis model. N Sets: Number of tested KEGG gene sets in which the gene is present. Presence (1) or absence (0) of gene in cholinergic (CH), glutamatergic (GL), dopaminergic (DO), GABAergic (GA), serotonergic (SE), LTP (LP) or LTD (LD) pathways. Z: Gene Z score. P: Gene association to schizophrenia (MAGMA gene p-value).

2.6 References

1. Lichtenstein P, Björk C, Hultman CM, Scolnick E, Sklar P, Sullivan PF. Recurrence risks for schizophrenia in a Swedish national cohort. Psychol Med 2006;36:1417–25.
2. Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. Nat Genet 2017;6:111.
3. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511(7510):421–7.
4. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 2018;199:441.
5. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nat Rev Genet 2016;17(6):353–64.
6. Duncan LE, Holmans PA, Lee PH, O'Dushlaine CT, Kirby AW, Smoller JW, et al. Pathway analyses implicate glial cells in schizophrenia. PLoS One 2014;9(2):e89441.
7. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA, et al. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol Psychiatry 2012;17(10):996–1006.
8. Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. Sci Rep 2017;7(1):12460.
9. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci 2015;18(2):199–209.
10. Hertzberg L, Katsel P, Roussos P, Haroutunian V, Domany E. Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia. Schizophr Res 2015;164(1-3):92–9.
11. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet 2011;7(1):e1001273.
12. GTEx Consortium, Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45(6):580–5.
13. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol 2015;11(4):e1004219.
14. Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics 2012;28(13):1797–9.
15. Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D, DIAGRAM Consortium, et al. Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. PLoS Genet 2010;6(8):e1001058.
16. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526(7571):68–74.
17. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet 2008;4(10):e1000214.
18. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res 2015;43(Database issue):D1049–56.
19. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27(12):1739–40.
20. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30.
21. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 2013;29(14):1830–1.
22. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4(1):7.

23. Warnes GR, Ben Bolker, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. [Internet]. [cited 2017 Nov 6];Available from: https://CRAN.R-project.org/package=gplots
24. Bernardinelli Y, Nikonenko I, Muller D. Structural plasticity: mechanisms and contribution to developmental psychiatric disorders. Front Neuroanat 2014;8(123).
25. Kahn RS, Sommer IE, Murray RM, Meyer-Lindenberg A, Weinberger DR, Cannon TD, et al. Schizophrenia. Nat Rev Dis Primers 2015;:15067.
26. Pocklington AJ, O'Donovan M, Owen MJ. The synapse in schizophrenia. Eur J Neurosci 2014;39:1059–67.
27. Kaalund SS, Newburn EN, Ye T, Tao R, Li C, Deep-Soboslay A, et al. Contrasting changes in DRD1 and DRD2 splice variant expression in schizophrenia and affective disorders, and associations with SNPs in postmortem brain. Mol Psychiatry 2014;19(12):1258–66.
28. Luykx JJ, Broersen JL, de Leeuw M. The DRD2 rs1076560 polymorphism and schizophrenia-related intermediate phenotypes: A systematic review and meta-analysis. Neurosci Biobehav Rev 2017;74, Part A:214–24.
29. Vink M, de Leeuw M, Luykx JJ, van Eijk KR, van den Munkhof HE, van Buuren M, et al. DRD2 Schizophrenia-Risk Allele Is Associated With Impaired Striatal Functioning in Unaffected Siblings of Schizophrenia Patients. Schizophr Bull 2016;42(3):843–50.
30. Salavati B, Rajji TK, Price R, Sun Y, Graff-Guerrero A, Daskalakis ZJ. Imaging-based neurochemistry in schizophrenia: a systematic review and implications for dysfunctional long-term potentiation. Schizophr Bull 2015;41:44–56.
31. Carruthers SP, Gurvich CT, Rossell SL. The muscarinic system, cognition and schizophrenia. Neurosci Biobehav Rev 2015;55:393–402.
32. Parikh V, Kutlu MG, Gould TJ. nAChR dysfunction as a common substrate for schizophrenia and comorbid nicotine addiction: Current trends and perspectives. Schizophr Res 2016;171(1-3):1–15.
33. Kahn RS, Keefe RS. Schizophrenia is a cognitive illness: time for a change in focus. JAMA Psychiatry 2013;70(10):1107–12.
34. Won H, la Torre-Ubieta de L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 2016;538(7626):523–7.
35. CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet 2017;49(1):27–35.
36. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature 2014;506(7487):179–84.
37. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 2012;17(2):142–53.
38. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506(7487):185–90.
39. Koh HY. Phospholipase C-beta1 and schizophrenia-related behaviors. Adv Biol Regul 2013;53(3):242–8.
40. Vasco Lo VR, Cardinale G, Polonia P. Deletion of PLCB1 gene in schizophrenia-affected patients. J Cell Mol Med 2012;16(4):844–51.
41. Forero DA, Herteleer L, De Zutter S, Norrback KF, Nilsson LG, Adolfsson R, et al. A network of synaptic genes associated with schizophrenia and bipolar disorder. Schizophr Res 2016;172(1-3):68–74.
42. Kyosseva SV. The role of the extracellular signal-regulated kinase pathway in cerebellar abnormalities in schizophrenia. Cerebellum 2004;3(2):94–9.
43. Yuan P, Zhou R, Wang Y, Li X, Li J, Chen G, et al. Altered levels of extracellular signal-regulated kinase signaling proteins in postmortem frontal cortex of individuals with mood disorders and schizophrenia. J Affect Disord 2010;124(1-2):164–9.

44. Liu L, Luo Y, Zhang G, Jin C, Zhou Z, Cheng Z, et al. The mRNA expression of DRD2, PI3KCB, and AKT1 in the blood of acute schizophrenia patients. Psychiatry research 2016;243:397–402.
45. van Winkel R. Family-based analysis of genetic variation underlying psychosis-inducing effects of cannabis: sibling analysis and proband follow-up. Arch Neurol 2011;68(2):148–57.
46. Pan B, Chen J, Lian J, Huang X-F, Deng C. Unique Effects of Acute Aripiprazole Treatment on the Dopamine D2 Receptor Downstream cAMP-PKA and Akt-GSK3β Signalling Pathways in Rats. PLoS One 2015;10(7):e0132722.
47. Kapfhamer D, Berger KH, Hopf FW, Seif T, Kharazia V, Bonci A, et al. Protein Phosphatase 2a and glycogen synthase kinase 3 signaling modulate prepulse inhibition of the acoustic startle response by altering cortical M-Type potassium channel activity. J Neurosci 2010;30(26):8830–40.
48. Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J Comp Neurol 2009;513(5):532–41.
49. Yin DM, Chen YJ, Sathyamurthy A, Xiong WC, Mei L. Synaptic dysfunction in schizophrenia. Adv Exp Med Biol 2012;970:493–516.
50. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph 2014;20(12):1983–92.
51. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 2011;12(1):35.
52. Binns D, Dimmer E, Huntley R, Barrell D, O’Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 2009;25(22):3045–6.


3

Chapter 3

Multivariate genome-wide analysis of stress-related quantitative phenotypes

Dick Schijven, Elbert Geuze, Christiaan H. Vinkers, Sara L. Pulit, Remmelt R. Schür, Marie Malgaz, Erwin Bekema, Jelena Medic, Kendrick E. van der Kust, Jan H. Veldink, Marco P. Boks, Eric Vermetten, Jurjen J. Luykx

Manuscript submitted

Abstract

Exposure to traumatic stress increases the odds of developing a broad range of psychiatric conditions. Genetic studies targeting stress-related quantitative phenotypes may shed light on mechanisms underlying vulnerability to psychopathology in the aftermath of stressful events. We applied a multivariate genome-wide association study (GWAS) to a unique military cohort (N = 583) in which we measured biochemical and behavioral phenotypes. The availability of pre- and post-deployment measurements allowed to capture changes in these phenotypes in response to stress. For genome-wide significant loci, we performed functional annotation, phenome-wide analysis and quasi-replication in PTSD case-control GWASs. We discovered one genetic variant reaching genome-wide significant association, surviving permutation and sensitivity analyses (rs10100651, $p = 9.9 \times 10^{-9}$). Functional annotation prioritized the genes *INTS8* and *TP53INP1*. A phenome-wide scan revealed a significant association of these same genes with sleeping problems, hypertension and subjective well-being. Finally, a targeted lookup revealed nominally significant association of rs10100651 in a PTSD case-control GWAS in the UK Biobank ($p = 0.02$). We provide comprehensive evidence from multiple resources hinting at a role of the highlighted genetic variant in the human stress response, marking the power of multivariate genome-wide analysis of quantitative measures in stress research. Future genetic and functional studies can target this locus to further assess its effects on stress mediation and its possible role in psychopathology or resilience.

3.1 Introduction

Military personnel report increased stress-related mental health symptoms after deployment, including mostly post-traumatic stress, depression, fatigue and anxiety [1,2]. Unambiguous clinical diagnosis of psychopathology following trauma is cumbersome, as supported by shifting diagnostic classification criteria of post-traumatic stress-disorder (PTSD) [3,4] and moderate to low inter-rater reliability in establishing a diagnosis of trauma-related disorders [5]. In studies investigating genetic mechanisms of disease, imprecise diagnostic ascertainment for case-control status may be circumvented by targeting phenotypes that correspond to more specific behavioral and biological mechanisms underlying stress-related psychopathology. Such phenotypes are reliably quantifiable, have reasonable heritability, lie closer to disease biology than clinical diagnoses, and generally show less measurement variability [6]; thus improving power for variant discovery. In addition, novel tools now allow for the integration of multiple layers of such phenotypic data into multivariate genome-wide analyses, increasing power relative to univariate analyses [7,8].

To find novel genetic loci associated with modulation of stress and psychopathology following trauma, we performed a multivariate genome-wide analysis of stress-related phenotypes (both biochemical and behavioral) in a military cohort. The collection of an extensive and unique set of pre- and post-deployment phenotypic measurements allowed us to carry out unprecedented multivariate analyses using post-deployment measurements corrected for pre-deployment measurements as independent variables and genome-wide SNP dosages as dependent variables (Figure 1), capturing the effect of deployment stress on these phenotypes. We report a genome-wide significant locus, perform several sensitivity analyses to show the robustness of the locus, assess the functional consequences this variant exerts on surrounding genes, and perform a phenome-wide analysis, highlighting association of functionally prioritized genes with other stress-related phenotypes. Finally, we discuss the potential future gains that may result from analyzing quantitative phenotypes in GWASs of post-traumatic stress.

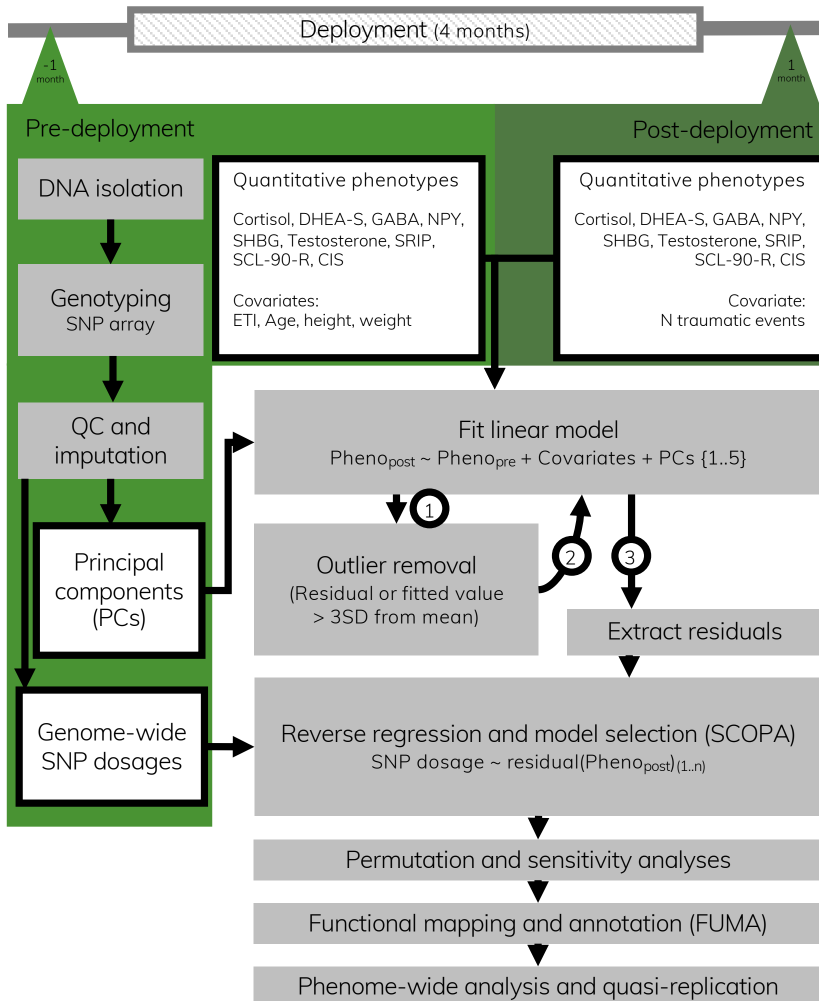


Figure 1. Analysis overview. Data on quantitative phenotypes were collected one month before (pre-deployment) and one month after (post-deployment) a 4-month deployment period. Additionally, DNA was isolated from whole blood and genotyped on a SNP array after which quality control (QC) and imputation procedures were completed. Principal components (PCs) and SNP dosages were extracted from imputed data. Residuals were calculated for each quantitative phenotype by fitting a linear model of the post-deployment measurement against the pre-deployment measurement, covariate phenotypes and the first five genetic PCs. Outliers in residuals or fitted values were removed. SNP dosage data and residuals were combined in reverse regression analysis using Software for Correlated Phenotype Analysis (SCOPA) and we followed up genome-wide significant hits with permutations, sensitivity analyses, functional annotation, phenome-wide analysis and quasi-replication in PTSD case-control GWASs. Abbreviations: DHEA-S, Dehydroepiandrosterone sulfate; GABA, Gamma-aminobutyric acid; NPY, Neuropeptide Y, SHBG, Sex hormone-binding globulin; SRIP, Self-Rating Inventory PTSD; SCL-90-R, Dutch version of the Revised 90-item symptom checklist; CIS, Checklist Individual Strength; ETI, Early Trauma Inventory; SD, Standard Deviation; PC, Principal Component.

3.2 Materials and methods

3.2.1 Participants and procedures

The data used for this study is part of an ongoing longitudinal study on stress-related mental health symptoms in Dutch military personnel. Individuals were deployed to Afghanistan during a four-month period between 2005 and 2008 as part of the International Security Assistance Force (ISAF). Approximately one month prior to deployment and one month after deployment, blood samples were collected and individuals completed questionnaires at the army base (Figure 1) [2]. Military personnel volunteering to take part in the study received verbal and written descriptions of the study and provided informed consent. The study was approved by the Institutional Review Board of the University Medical Center Utrecht (Utrecht, the Netherlands). We adhered to the ethical statements described in the declaration of Helsinki [9].

3.2.2 Phenotype collection and measurements

Blood samples were collected between 8 a.m. and 11.30 a.m. in EDTA vacutainers that were cooled on ice immediately after collection. Samples were centrifuged at 3,500 rpm for 12 minutes at 4 °C and stored at -80 °C until measurements were performed. Plasma concentrations of dehydroepiandrosterone sulfate (DHEA-S), a metabolite of the anti-glucocorticoid hormone DHEA [10], were measured using an electrochemiluminescence immunoassay on the Modular E170 (Roche Diagnostics GmbH, Mannheim, Germany). The lower limit of detection was 0.05 μmol/l with an inter-assay variation < 4.4% in the range of 1.7 – 12 μmol/l. Other measurements included cortisol, the most important human glucocorticoid [11]; gamma-aminobutyric acid (GABA), a modulator of hypothalamic-pituitary-adrenal (HPA)-axis activity [12]; neuropeptide Y (NPY), a peptide neurotransmitter and modulator of the stress response [13]; testosterone, a product of the hypothalamic-pituitary-gonadal (HPG)-axis and involved in psychosocial and behavioral phenotypes [14]; and sex hormone-binding globulin (SHBG), a protein produced by the liver regulating bioavailable testosterone [14]. These were quantified in plasma as described in the referenced manuscripts.

Behavioral phenotypes were measured using surveys and clinically-implemented symptom checklists. We assessed subject mental health using the Dutch revised Symptom Checklist (SCL-90-R) [15], evaluated symptoms of PTSD

using the Dutch Self-Rating Inventory for PTSD (SRIP) [16,17], measured fatigue using the Checklist Individual Strength (CIS) [18], gathered data on combat-related stressors and potentially traumatic experiences using a 19-item checklist [2], and finally assessed exposure to trauma before the age of 18 using the Early Trauma Inventory (ETI) [19] (Supplementary Table 1).

3.2.3 Genotyping, quality control, imputation and calculation of genetic relationship matrix

DNA from a total of 1,015 individuals was isolated from whole blood at University Medical Center Utrecht (Utrecht, the Netherlands) and shipped to IoPPN Genomics & Biomarker Core Facility of King's College London for genotyping on the Illumina Human Omniexpress-24 v1.1 platform (Illumina, San Diego, CA, USA). This platform contains a total number of 713,040 genetic markers. Genotypes were called using Illumina GenomeStudio software without any filtering presets. Subsequently, we applied quality control (QC) to the genotype data using PLINK version 1.90b3z (22-11-2015; www.cog-genomics.org/plink/1.9) [20]. We removed samples with a genotype call rate < 0.95, discordant genetic sex and reported sex in phenotype data, and an excess heterozygosity rate (heterozygosity > 3SD from the mean) or excess homozygosity (heterozygosity < 3SD from the mean). Additionally, we identified pairs of individuals with a relatedness coefficient (PLINK PIHAT) > 0.1 and randomly removed one individual from the pair. Next, we performed SNP-level QC: we removed non-autosomal genetic markers, SNPs with a call rate < 0.95, SNPs out of Hardy-Weinberg Equilibrium (HWE, p-value < 1×10^{-6}) and SNPs with a strand-ambiguous A/T or C/G genotype. In preparation for imputation, we removed SNPs that were not present in the reference panel for imputation or had a difference in minor allele frequency (MAF) > 0.15 compared to the reference panel.

We imputed the genetic data on the Michigan Imputation Server [21], using the Haplotype Reference Consortium (HRC) release 1.1 as a reference panel [22]. The Michigan Imputation Server phases the data before imputation using the Eagle2 algorithm [23]. Imputation was performed with minimac3 in a 500 kilobase (Kb) window. We applied additional quality control to the imputed data (post-imputation QC): we removed SNPs with a MAF < 0.05, imputation R^2 < 0.3 and a MAF difference from the reference panel > 0.15. We used

genotype dosages in our analyses, unless stated otherwise.

We generated a genetic relationship matrix (GRM) using Genome-wide Complex Trait Analysis (GCTA) software (version 1.24.4) [24], and subsequently calculated 100 principal components (PCs) from this GRM. The first five PCs were used as covariates in analyses to capture genetic variation due to population stratification.

3.2.4 Multi-phenotype association analysis

Prior to the association analysis, we excluded individuals with non-male sex (due to the low number of females in the cohort and sex differences in hormone levels), missing data for analysis covariates and without a completed pre- and post-deployment measurement for at least one stress-related phenotype (cortisol, DHEA-S, GABA, NPY, testosterone, SHBG; and total scores for SRIP, SCL-90-R and CIS). To capture genetic association with the effect of deployment stress, we fitted post-deployment measurements on pre-deployment values of the included phenotypes using R version 3.3.3 (www.r-project.org). We additionally included age, ETI total score, number of experienced potentially traumatic events during deployment and five PCs based on the genetic data as covariates in the regression model and additionally included pre-deployment height and weight as covariates for the biochemical phenotypes. We removed samples when the residual or fitted value deviated more than 3 SD from the mean, then fitted the linear model again for the remaining individuals, extracted residuals, and used these as input for the multiple-phenotype association analysis in Software for Correlated Phenotype Analysis (SCOPA) [7]. In SCOPA we applied reverse regression for each SNP, with genotype dosage as the dependent variable and post-deployment phenotype residuals as the independent variables. Additionally, Bayesian Information Criterion (BIC) for optimal model selection was applied to variants reaching the established stringent threshold for genome-wide significance ($p < 5 \times 10^{-8}$), where the combination of phenotypes with the lowest BIC resulted in the optimal model.
For genome-wide significant hits, we ran an additional association analysis in a ± 2 Mb window around the top variant testing all SNPs with the optimal model (lowest BIC) of the genome-wide significant variant.

3.2.5 Permutation and sensitivity analyses of genome-wide significant hits

We performed several sensitivity analyses on lead SNPs of loci that reached genome-wide significance in the main multivariate analysis. First, we performed $n = 1 / p$ permutations of sample labels and subsequent SCOPA regression using the optimal model for the corresponding genome-wide significant variant, where p is the nominal association p-value of the variant. We calculated an empirical p-value for each top variant using $p_{emperical} = (s + 1) / (n + 1)$, where s is the number of times the permuted p-value was lower than the nominal p-value and n the total number of permutations performed. Second, we performed the same analysis pipeline using log-transformed questionnaire data to reduce skewness in the distribution of residuals. Finally, we substituted the questionnaire total scores with the scores of the subscales, as these pinpoint more specific phenotypic domains, and repeated the analysis pipeline. For each of the sensitivity analyses, we applied the same threshold for genome-wide significance ($p < 5 \times 10^{-8}$).

3.2.6 Functional annotation

We used Functional Mapping and Annotation (FUMA) for annotation of genome-wide significant loci with expression quantitative trait locus (eQTL) effects and chromatin interactions [25]. In FUMA, eQTL analysis included Genotype-Tissue Expression (GTEx) (version 6 and 7) [26,27], blood eQTL browser [28], BIOS QTL browser [29], BRAINEAC [30], MuTHER [31], xQTLServer [32] and the CommonMind Consortium [33] datasets. Chromatin interaction data was derived from the GSE87112 database [34].

3.2.7 Phenome-wide analysis and quasi-replication

The unique longitudinal design and phenotype collection in the cohort precluded us from finding a dataset with identical phenotypes to replicate genome-wide significant loci. To assess the role of genome-wide significant loci and functionally prioritized genes in other phenotypes, we first performed a hypothesis-generating phenome-wide scan using the GWAS atlas (atlas.ctglab.nl) [35]. The GWAS atlas contains SNP- and gene-level association results of publicly available GWASs and UK Biobank traits with at least 50,000 individuals and more than 10,000 cases, totaling 2,824 unique phenotypes (accessed 29-11-2018). Second, to test how our findings reflect genetic risk for PTSD, the main diagnosis related to trauma-induced psychopathology, we

performed a quasi-replication of any genome-wide significant loci in the European subset of the largest published case-control GWAS meta-analysis on PTSD (2,424 cases; 7,113 controls) [36], and in a case-control GWAS of self-reported PTSD from the UK Biobank (266 cases; 360,875, accessed August 2018) [37,38]. The first dataset was also included in the phenome-wide analysis, but the latter was not due to the low number (< 10,000) of cases. From the UK Biobank PTSD GWAS summary statistics, we first removed SNPs with an expected minor allele count < 25 or a minor allele frequency < 0.05 as these are likely to inflate the number of false positives given the low number of cases and disbalance in case-control ratio.

3.3 Results

3.3.1 Descriptive statistics

From our initial study population after genotype QC (n = 963), we excluded the following samples: non-male sex (n = 86), missing reported age (n = 5), missing reported number of traumatic events during deployment (n = 292), missing reported ETI total score (n = 91), missing reported pre-deployment height (n = 47) and missing reported pre-deployment weight (n = 88), the latter two being only relevant for biochemical phenotypes.

The mean pre-deployment age of the included participants was 29.27 (SD 9.40) with an exposure to a mean of 4.55 (SD 3.23) potentially traumatic events during deployment. Mean concentrations of all biochemical compounds and means of questionnaire total scores had increased through deployment (Table 1). After calculating residuals of post-deployment measurements and removal of outliers, 560 individuals with at least one complete pre- and post-deployment measurement of a single phenotype remained. 293 participants had non-missing data for all phenotypes (Supplementary Figure 1). Phenotypic correlations calculated in these 293 individuals revealed positive relationships between behavioral questionnaires; DHEA-S, testosterone and SHBG; and DHEA-S and cortisol, and a negative relationship between NPY and SHBG (Supplementary Figure 2).

3.3.2 Multivariate association testing

After quality control and imputation, the dataset comprised > 5 million variants

Table 1. Descriptive statistics of male individuals after application of exclusion criteria.

		Pre-deployment			Post-deployment		
	n	Mean [range]	SD	Median	Mean [range]	SD	Median
General statistics (covariates) (total N = 583)[1]							
Age at start of deployment (years)	583	29.27 [18-60]	9.40	26	-	-	-
ETI total score	583	3.46 [0-17]	3.01	3	-	-	-
Potential traumatic events during deployment	583	-	-	-	4.55 [0-14]	3.23	4
Height at start of deployment (cm)	575	183 [165-202]	6.53	183	-	-	-
Weight at start of deployment (kg)	556	83 [58-130]	10	83	-	-	-
Biochemical data (total N = 556)[2]							
Cortisol (nmol/l)	524	439.83 [122-1412]	141.35	433.5	449.30 [143-917]	135.99	443.5
DHEA-S (µmol/l)	520	10.25 [1.48-22.12]	3.49	9.99	10.33 [1.92-25.97]	3.84	10.02
GABA (nmol/l)	521	120.08 [70-201]	21.93	118	124.74 [75-217]	23.08	121
NPY (ng/ml)	522	41.40 [15-236.9]	17.55	38.05	43.15 [10.8-160.5]	18.29	40.5
SHBG (nmol/l)	520	19.07 [7-58]	6.39	18	19.80 [4-50]	6.54	19
Testosterone (nmol/l)	520	17.87 [0.65-39]	5.95	18	19.16 [4.9-280]	13.61	18
Questionnaire data (total N = 583)[1]							
SRIP total score	431	26.87 [22-57]	4.98	25	27.92 [22-57]	6.08	26
SCL-90-R total score	496	102.40 [90-209]	14.57	98	103.30 [90-194.65]	17.40	97
CIS total score	519	45.91 [20-101]	17.52	43	49.32 [20-111]	21.45	44

[1] After exclusion of individuals with missing data on age, ETI and potential traumatic event count.

[2] After exclusion of individuals with missing data on age, ETI, potential traumatic event count, pre-deployment height and pre-deployment weight.

Sample size column may not add up to total sample size per data category due to missing values.

Abbreviations: n, number of individuals with complete data; SD, Standard Deviation; ETI, Early Trauma Inventory; DHEA-S, Dehydroepiandrosterone sulfate; GABA, Gamma-aminobutyric acid; NPY, Neuropeptide Y; SHBG, Sex hormone-binding globulin; SRIP, Self-Rating Inventory PTSD; SCL-90-R, Dutch version of the Revised 90-item symptom checklist; CIS, Checklist individual strength.

(Supplementary Table 2, Supplementary Table 3). Samples were primarily of European descent, with the vast majority (± 91%) lying < 6 SD from the HapMap3 European populations along PCs one and two (Supplementary Figure 3).

Multiple-phenotype reverse regression analysis revealed two SNPs passing the threshold for genome-wide significance ($p < 5 \times 10^{-8}$) under optimal models composed of several intermediate phenotypes (Figure 2A, Table 2). The first SNP maps to an intron of the Integrator Complex Subunit 8 gene (*INTS8*) on chromosome 8 (rs10100651, $p = 9.9 \times 10^{-9}$) in a regression model including DHEA-S, GABA, NPY, SHBG, SCL-90-R and SRIP data (BIC = 94.44) (Table 2, Supplementary Table 4). DHEA-S was the strongest contributor to the association (beta = 0.12, se = 0.023) (Supplementary Table 5, Supplementary Figure 4). This SNP had a G/C genotype with a MAF of 0.46 (C-allele). To ensure no strand inconsistencies influenced association at this G/C variant with high MAF, we calculated haplotype frequencies including surrounding variants in high linkage disequilibrium (LD $R^2 > 0.7$) with rs10100651, of which one variant was directly genotyped and the others were imputed, using HaploView software (version 4.2) (Supplementary Table 6) [39]. We found five different haplotypes, the most common of which were comprised of either only major alleles (0.51) for all SNPs, or minor alleles for all SNPs (0.39). We also calculated haplotypes in the European subset of 1000 Genomes Phase 3 and compared them to haplotype frequencies in our cohort [40]. None of the identified haplotypes in our data had > 7% frequency difference when compared to haplotype frequencies in 1000 Genomes. Overall, these results suggested no strand inconsistencies at this locus in our dataset.

The second SNP we discovered in the multi-phenotype regression was an intronic variant of the MAM Domain Containing Glycosylphosphatidylinositol Anchor 2 gene (*MDGA2*) on chromosome 14 (rs4900759, $p = 1.9 \times 10^{-8}$). The optimal model included total scores of SRIP and SCL-90-R questionnaires as phenotypes (BIC = -15.22) (Table 2, Supplementary Table 4), with SRIP being the strongest contributor to the association (beta = -0.06, se = 0.009) (Supplementary Table 5).
To assess statistical credibility of the discovered loci, we performed permutation of sample labels followed by reverse regression in SCOPA with the corresponding

3

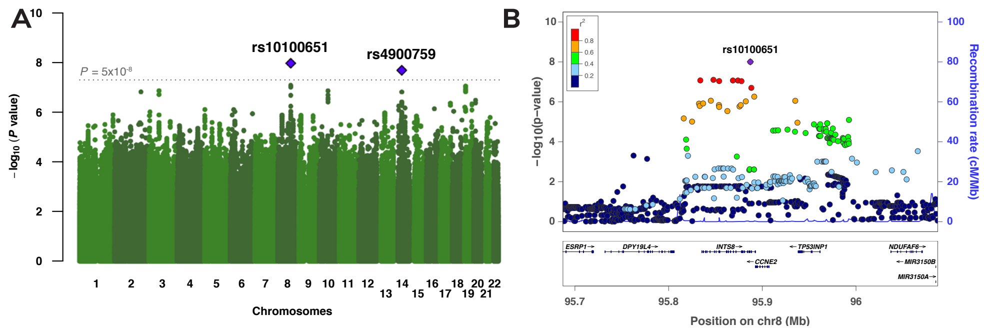


Figure 2. Multivariate GWAS result. A) Manhattan plot showing all tested variants with their relative position across the chromosomes (x-axis) and their association with post-deployment quantitative phenotype residuals at their optimal model as selected through BIC in SCOPA (y-axis). Genome-wide significance threshold (dotted line) is $p < 5 \times 10^{-8}$. Diamonds indicate variants reaching genome-wide significance. **B)** Regional association plot of rs10100651 on chromosome 8. All SNPs in this plot were tested with the optimal model for lead SNP rs10100651 (including DHEA-S, GABA, NPY, SHBG, SRIP and SCL-90-R). Colors indicate LD (R^2) in the European population of surrounding variants relative to rs10100651.

optimal model for the top SNPs of the discovered loci. Permutations at rs10100651 resulted in an empirical p-value of $p = 2.0 \times 10^{-8}$. Permutation analysis for rs4900759 resulted in an empirical p-value of $p = 3.8 \times 10^{-8}$ (Table 2). To reduce skewness in the residuals of questionnaire data (Supplementary Figure 1), we then log-transformed questionnaire total scores, re-calculated residuals and reran the analysis pipeline for the two genome-wide significant hits. The association signal at rs10100651 remained significant at $p = 2.9 \times 10^{-8}$, with the optimal model identified through BIC remaining similar to the main analysis (DHEA-S, GABA, NPY, SHBG, SCL-90-R and SRIP), whereas the association at rs4900759 dropped below significance ($p = 1.5 \times 10^{-5}$) and the model differed from the optimal model in the main analysis (GABA, NPY, SCL-90-R, SRIP) (Supplementary Table 5).

As a final sensitivity analysis, questionnaire total scores were replaced by subscale scores and the analysis pipeline for the genome-wide significant SNPs was rerun. For rs10100651, the p-value had further decreased ($p = 3.2 \times 10^{-9}$) with an optimal model composed of the same biochemical phenotypes as in the original analysis (DHEA-S, GABA, NPY and SHBG). Subscales of the same questionnaires as in the primary analysis were present in the optimal model (SRIP avoidance, SRIP re-experiencing and SCL-90-R sleeping problems). For rs4900759, again the p-value increased to non-significance ($p = 1.6 \times 10^{-6}$) with a different optimal model (Supplementary Table 5). Therefore, all these analyses

Table 2. Statistics of genome-wide significant hits in multivariate GWAS.

SNP	Effect (other) allele	HWE	MAF	R^2	N	null Log Likelihood	Likelihood Ratio	P-value	BIC	Model	P_{perm}
rs10100651	C (G)	0.30	0.46	1.00	307	-51.37	48.39	9.9×10^{-9}	94.44	DHEA-S GABA NPY SHBG SRIP SCL-90-R	2.0×10^{-8}
rs4900759	T (C)	0.23	0.26	1.00	381	-1.25	35.55	1.9×10^{-8}	-15.22	SRIP SCL-90-R	3.8×10^{-8}

Abbreviations: SNP, Single-nucleotide polymorphism; HWE, p-value of Hardy-Weinberg Equilibrium test; MAF, Minor allele frequency; R^2, Imputation quality; N, sample size of the model; BIC, Bayesian Information Criterion; P_{perm}, Empirical p-value calculated from permutation analysis; DHEA-S, Dehydroepiandrosterone sulfate; GABA, Gamma-aminobutyric acid; NPY, Neuropeptide Y, SHBG, Sex hormone-binding globulin; SRIP, Self-Rating Inventory PTSD; SCL-90-R, Dutch version of the Revised 90-item symptom checklist.

confirmed the robustness of the chromosome 8 locus (rs10100651) to permutations and phenotype rescaling, while diminishing the evidence for the signal at chromosome 14 (rs4900759). For rs10100651, we analyzed all variants in a ±2 Mb window surrounding this SNP using the optimal model from the primary analysis to highlight association and LD structure at this locus (Figure 2B).

3.3.3 Functional annotation, phenome-wide analysis and quasi-replication in PTSD

Functional annotation of rs10100651 using eQTL and chromatin interaction data implemented in FUMA revealed both significant eQTL- and chromatin interaction of (the genomic region including) rs10100651 on two surrounding genes (*INTS8*; *TP53INP1*, Tumor Protein P53 Inducible Nuclear Protein 1) and either eQTL or chromatin interaction effects on other genes (Supplementary Table 7, Supplementary Table 8, Supplementary Figure 5).

Hypothesis-free phenome-wide analysis of the functionally prioritized genes *INTS8* and *TP53INP1* was performed using the GWAS atlas. Bonferroni correction of the p-value threshold for significance was applied to two genes and 2,824 unique traits assessed ($p = 8.9 \times 10^{-6}$). The analysis revealed significant association of these genes with multiple phenotypes (Figure 3, Supplementary Table 9, Supplementary Table 10), including: insomnia

(*TP53INP1*, p = 6.7 x 10^{-8}) and sleeplessness (*TP53INP1*, p = 3.3 x 10^{-6}), subjective well-being and general happiness (*TP53INP1*, p = 7.8 x 10^{-7}), hypertension (*TP53INP1*, p= 1.3 x 10^{-7}; *INTS8*, p = 3.42 x 10^{-7}) and blood pressure (*INTS8*, p = 3.3 x 10^{-8}). A specific scan of rs10100651 revealed no phenotypes passing the significance threshold, and this variant was only present in 52 GWAS results.

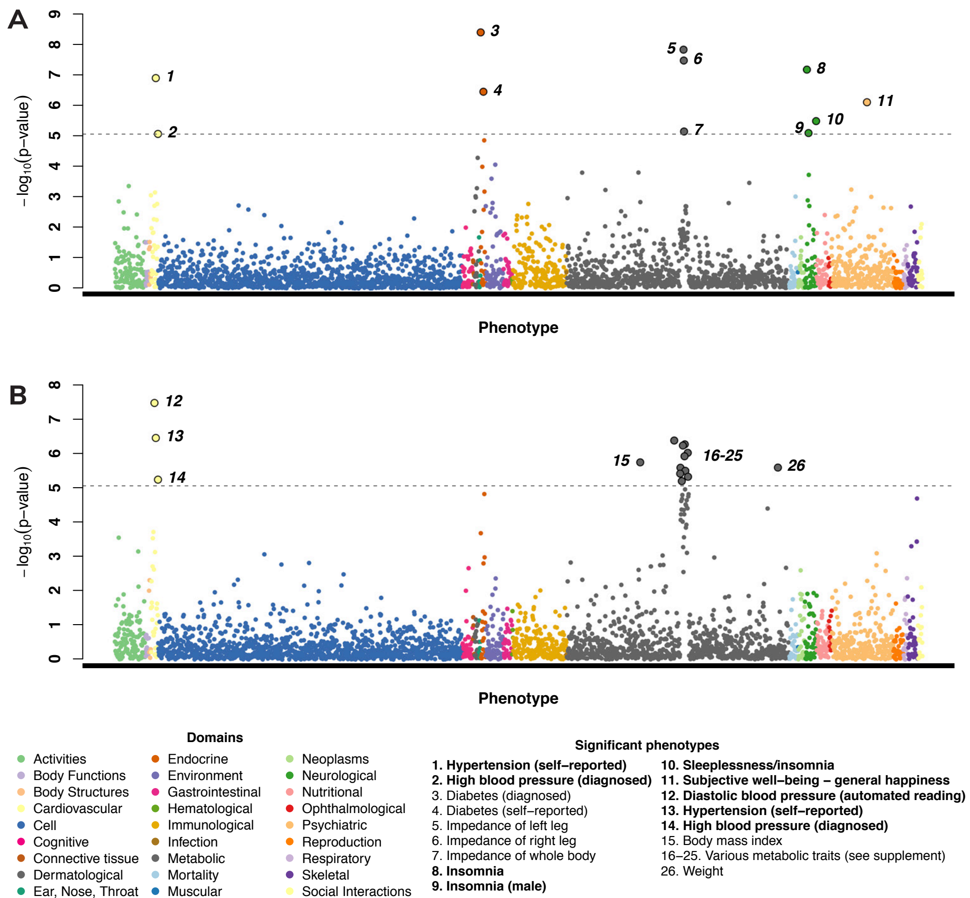


Figure 3. Phenome-wide analysis of *TP53INP1* and *INTS8*. Association -$\log_{10}$-converted p-values (y-axis) of **A)** *TP53INP1* and **B)** *INTS8* are shown for 2,824 phenotypes across 27 domains (indexed on the x-axis). Phenotypes reaching significance ($p < 8.9 \times 10^{-6}$, dashed line) are numbered and correspond to the significant phenotypes listed (stress-related phenotypes are highlighted). Association results for all phenotypes reaching significance (indicated by the horizontal line) are shown in Supplementary Tables 9 and 10. The results of this analysis were obtained from the GWAS atlas (atlas.ctglab.nl) [35].

We additionally performed a targeted lookup of rs10100651 in existing PTSD case-control GWASs. rs10100651 (MAF = 0.48, C-allele) passed SNP filtering for low MAF and minor allele count in the UK Biobank self-reported PTSD GWAS data (Supplementary Figure 6A). This SNP showed nominally significant association and directional consistency with DHEA-S and SRIP total score in our cohort (C-allele, beta = 1.5×10^{-4}, se = 6.4×10^{-5}, p = 0.02) (Supplementary Figure 6B). The SNP was contained in a locus with rs12550765 as a lead SNP (LD with rs10100651 $R^2 = 0.47$ in the British population, $R^2 = 0.34$ in the European population, association p = 6.7×10^{-4}), a SNP also showing nominally significant association in our multi-phenotype GWAS (p = 2.1×10^{-3}) when tested with the same set of phenotypes as the optimal model for rs10100651. In the largest PTSD case-control GWAS meta-analysis, association at rs10100651 did not reach significance (p = 0.18 in a meta-analysis limited to male subjects of European ancestry).

3

3.4 Discussion

By performing the first multivariate GWAS on stress-related longitudinally collected phenotypes, we discovered one locus reaching genome-wide significance. The association at this locus remained significant after permutation and sensitivity analyses, the lead SNP exerts functional effects on surrounding genes that were in turn associated with multiple stress-related phenotypes, and the locus quasi-replicated in a case-control GWAS of self-reported PTSD.

The optimal regression model for association at the highlighted locus (rs10100651) included four biochemical phenotypes (DHEA-S, GABA, NPY, SHBG) and two behavioral questionnaires (SRIP and SCL-90-R) with the strongest additive genetic effects at DHEA-S (positive), SHBG (negative) and GABA (negative; Supplementary Figure 4). This combination of associations reflects genotype-dependent differences in the change of these hormone levels throughout deployment and a possible effect of this locus on stress responsiveness. DHEA and DHEA-S are known to be increased in veterans suffering from PTSD as well as in trauma-exposed controls compared to non-trauma exposed controls, and this relates to symptom improvement [10,41]. In addition, DHEA has antioxidant effects and is a possible neuroprotective compound in stressed individuals [42-44]. The role of SHBG in stress and PTSD is

largely unknown, although it is well established that SHBG regulates the amount of bioavailable testosterone. The close association between SHBG and testosterone is shown by the positive correlation in post-deployment residuals in our cohort (Supplementary Figure 2). The negative direction of effect at rs10100651 possibly indicates less sensitivity to changes in testosterone levels at this variant, which has been associated with lower post-traumatic stress symptoms after deployment. In contrast, individuals with larger pre-post deployment changes of testosterone report higher levels of PTSD symptoms [14]. Although increased GABA was not found to be a strong predictor of psychopathology right after deployment in this cohort, an increase of GABA one month and six months post-deployment is associated with more mental health problems, suggesting stronger increase in GABA levels as a risk factor for post-trauma psychopathology [12]. The negative effect direction at the identified locus hints at a protective role for this variant in stress-related GABA regulation. The role of GABA in stress-induced psychopathology is not completely clear, as illustrated by lower plasma GABA levels in individuals who meet criteria for PTSD immediately after deployment [45]. Overall, the effect directions for these hormones at this SNP are in agreement: the positive effect at deployment-induced changes in DHEA-S (a stress-protective factor) and the negative effects at SHBG (in relation to testosterone a stress-inducing factor when levels strongly change) and GABA (in this cohort a stress-inducing factor upon strong increase) show the possible relevance of this variant as a factor involved in stress and resilience. Furthermore, different effect directions at GABA and DHEA-S are of interest as DHEA-S is an antagonist of the $GABA_A$ receptor, suggesting opposing functions in the maintenance of the inhibitory-excitatory balance with regard to stress for these hormones [46].

The lead SNP, rs10100651, showed significant eQTL effects and/or chromatin interactions with surrounding genes and in particular *INTS8* and *TP53INP1* (Supplementary Table 7, Supplementary Table 8). These genes pose several tantalizing biological leads: *INTS8* codes for part of the integrator complex, has been linked to human brain development [47]; and *TP53INP1* is involved in the regulation of cellular homeostasis in immune cell compartments and operates as an antioxidant upon exposure to stress [48]. Both oxidative stress and accelerated ageing of neurons are proposed mechanisms in PTSD [49]. Furthermore, phenome-wide analysis highlighted the SNP-based association

of these genes with several phenotypes that closely relate to stress. First, post-traumatic stress has been associated with high blood pressure and increased cardiovascular disease [50]. Second, sleeping problems are one of the symptoms defining the diagnostic classification criteria for PTSD [4], and disturbed sleep has been linked to severity of mental health symptoms after trauma [51]. Third, subjective wellbeing is negatively correlated to PTSD and depressive symptoms and increases upon treatment of these symptoms [52]. These identified phenotypes are furthermore correlated to each other, explaining the pleiotropic effect at these genes. eQTLs and chromatin interactions are ubiquitous throughout the genome, so it is important to note that these functional annotations do not necessarily highlight the likely causal genes involved in these phenotypes.

Although the highlighted locus is likely broadly associated to stress, we assessed its association to PTSD, which is the main diagnosis in relation to psychopathology after trauma. In the UK Biobank case-control GWAS on self-reported PTSD [38], rs10100651 reached nominal significance (Supplementary Figure 6), but the locus did not show association in the largest published meta-analysis. The replication in one out of two datasets might reflect a difference in definition of PTSD susceptibility between the replication sets. In the cohorts included in the PGC meta-analysis, PTSD case-control status was ascertained using DSM-IV based methods ranging from questionnaires to clinical interviews [36], while in the UK Biobank GWAS a self-reported case-control status independent of DSM criteria was used (which is line with the self-reported longitudinal data collected here). In addition, continuous stress-related phenotypes may increase statistical power in a GWAS, especially in a multivariate context [7,8].

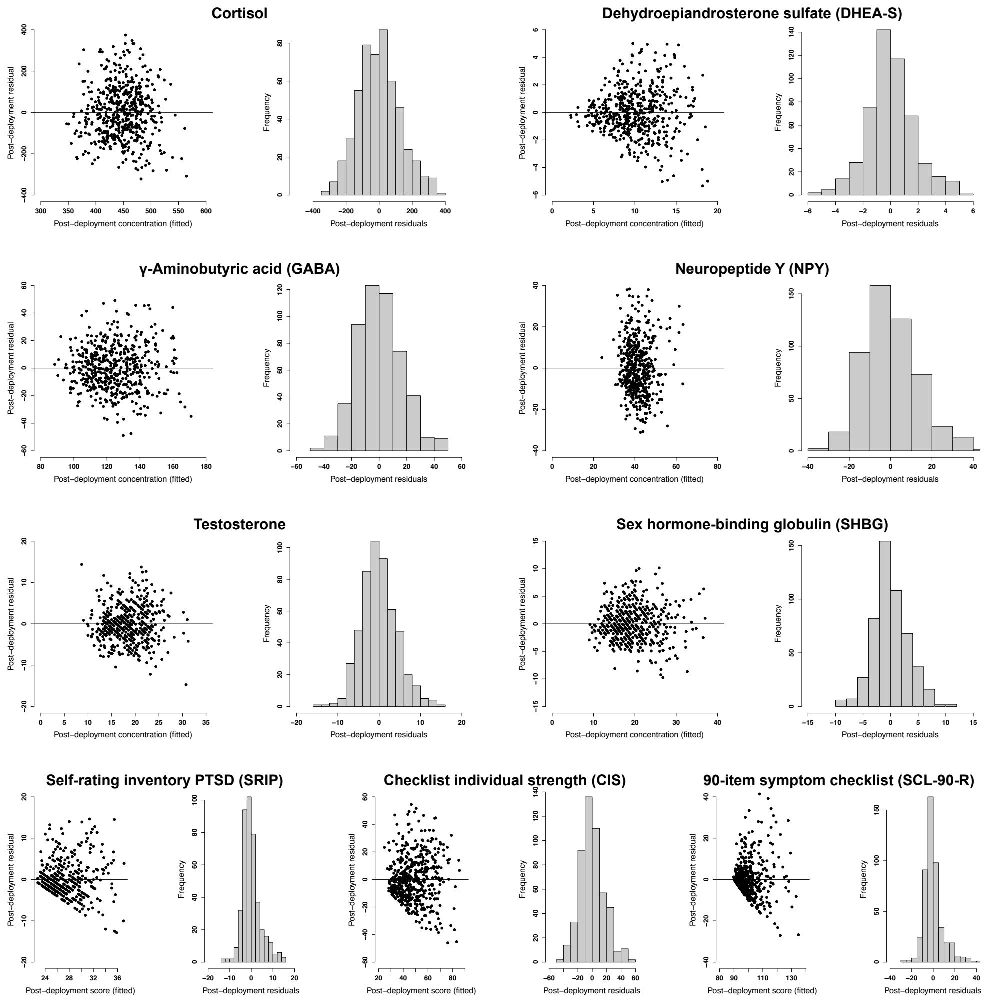
The number of GWASs using quantitative phenotypes underlying post-traumatic mental health conditions has so far been limited, but some associations have been reported [53,54]. Although ours is the first to implement multiple longitudinally measured phenotypes into a multivariate GWAS, our study also has limitations. First, our sample size was small which is in part due to data incompleteness for several participants. This a common problem in longitudinal studies dissecting a range of behavioral and biological phenotypes. On the other hand, our model of stress susceptibility yields substantial power to

identify loci underlying biological mechanisms owing to the quantitative nature of the data. Second, other well-established stress-related intermediate phenotypes, such as brain imaging parameters, were not available in our dataset but may in the future add an extra layer of information to these multiple phenotype genetic studies [55]. Third, our analysis included only males. Given the strong association with biochemical phenotypes of which the plasma concentrations are highly variable between males and females, this association might be specific to males. Finally, we lacked a replication cohort with similar phenotypes, precluding generalization of our findings to other (military) cohorts.

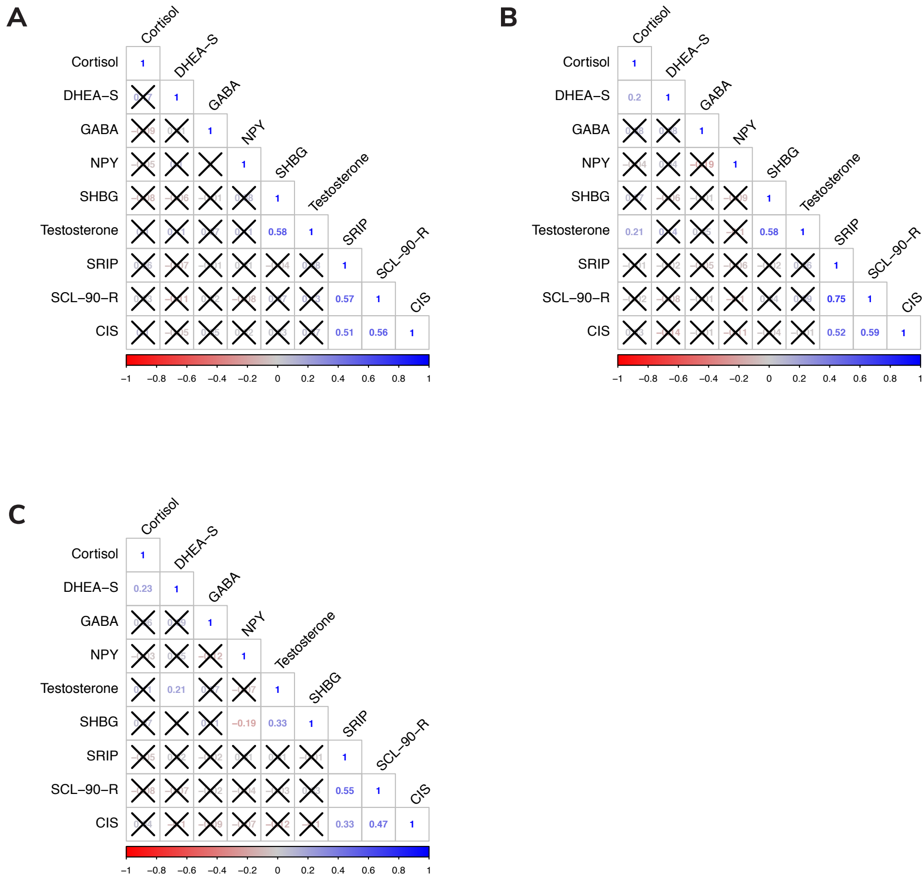
In conclusion, we have demonstrated the potential of targeting multiple quantitative phenotypes underlying stress sensitivity and highlight a credible genetic locus that may play a role in the modulation of the stress response upon exposure to trauma. This locus may be targeted in future downstream analyses, including fine-mapping and analyses of additional epigenomic annotations. As our understanding of the biological substrates involved in stress-related psychopathology increases, so do large-scale initiatives to collect genetic and quantitative phenotype data on these conditions [56,57]. Future studies using multivariate GWAS approaches may further refine heritability and genetic risk factors underlying susceptibility to adverse effects of traumatic stress.

3.5 Supplementary information

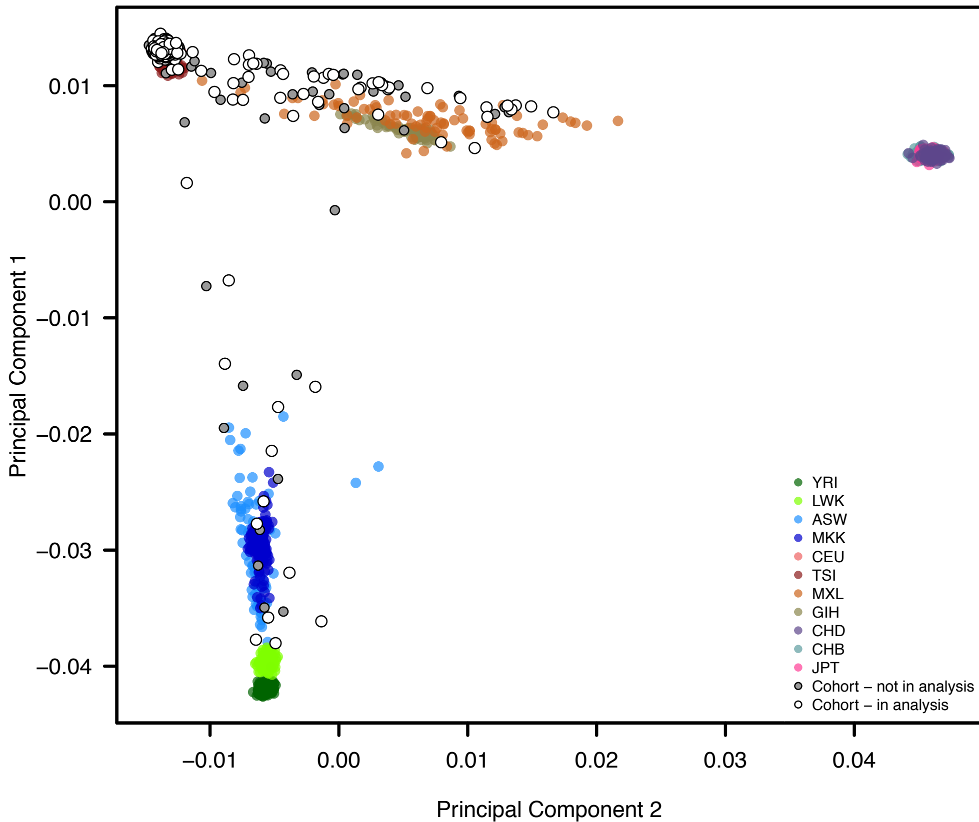
3.5.1 Supplementary figures



Supplementary Figure 1. Residuals used as input in multiple-phenotype GWAS. For each quantitative phenotype, the post-deployment measurement was fitted in a linear model against the pre-deployment measurement and relevant covariates (age, ETI total score, number of experienced potentially traumatic events during deployment, five genetic principal components, and additionally pre-deployment height and weight for biochemical phenotypes). Residuals were extracted from these models and used as input for multiple-phenotype GWAS. Residuals plotted against fitted values (scatter plot in the left panel) and the distribution of residuals (histogram in the right panel) are shown for each phenotype.

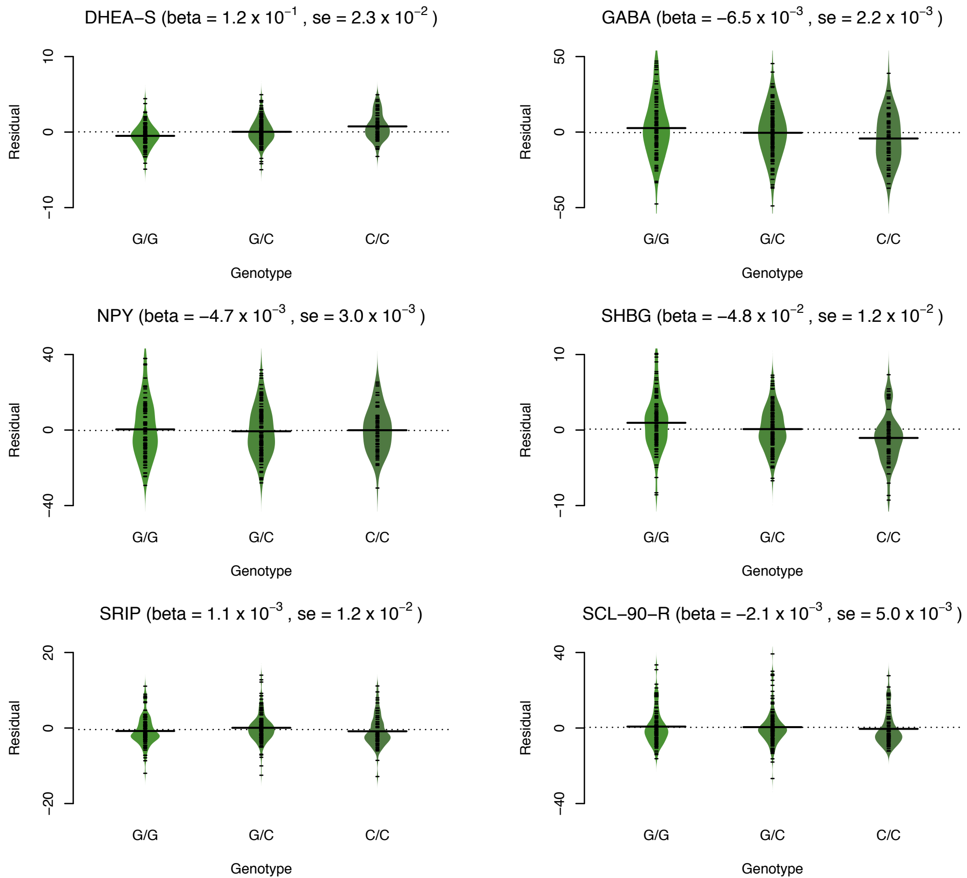


Supplementary Figure 2. Correlations between input phenotypes. Correlation matrices indicate positive (blue shades) and negative (red shades) correlations between **A)** pre-deployment phenotype measurements, **B)** post-deployment phenotype measurements and **C)** residuals of post-deployment measurements (input for multivariate GWAS) from a linear model with the pre-deployment measurement and covariates. Correlations are based on 293 individuals that had non-missing data for all included phenotypes. For significant correlations ($p < 0.05/36$ correlations tested) the value of the correlation is shown, non-significant values are crossed out.

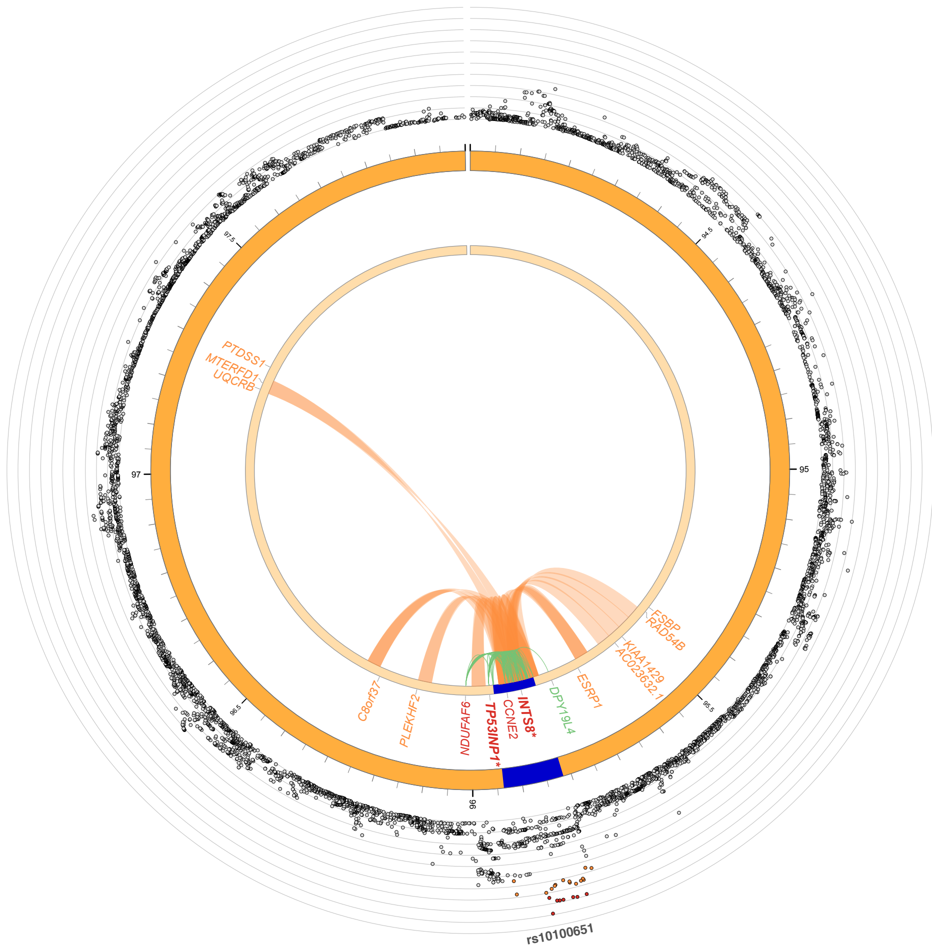


Supplementary Figure 3. Population structure of the cohort. The first two principal components based on the genetic data of the cohort are plotted on a HapMap background. Cohort individuals that passed genetic data quality control but did not pass inclusion criteria for the analysis (gray) and cohort individuals that passed QC and meet inclusion criteria for the analysis (white) are indicated separately. Included HapMap3 populations are Yoruba in Ibadan, Nigeria (YRI); Luhya in Weyube, Kenya (LWK); African ancestry in Southwest USA (ASW); Maasai in Kinyawa, Kenya (MKK); Utah residents with Northern and Western European ancestry (CEU); Toscani in Italy (TSI); Mexican ancestry in Los Angeles, California (MXL); Gujarati Indians in Houston, Texas (GIH); Chinese in Metropolitan Denver, Colorado (CHD); Han Chinese in Beijing, China (CHB); Japanese in Tokyo, Japan (JPT) [58].

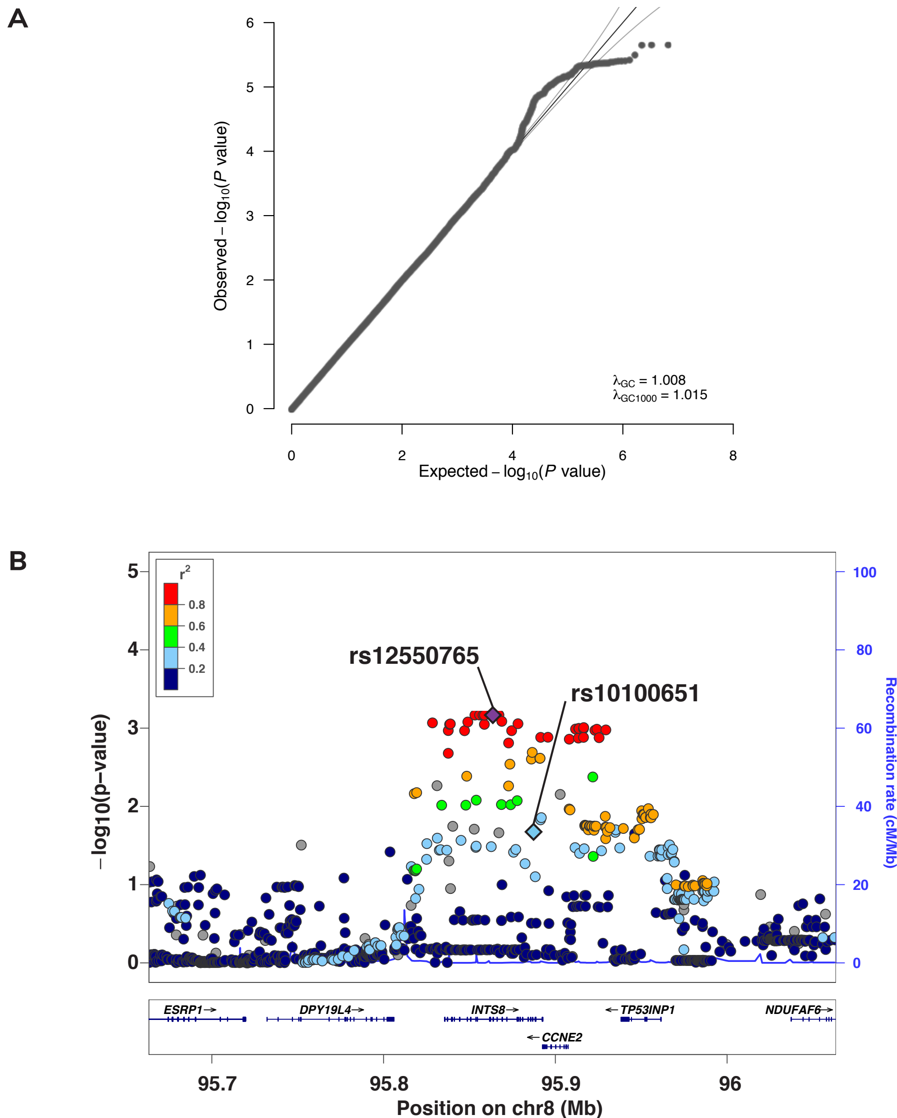
3



Supplementary Figure 4. Distribution of intermediate phenotype residuals stratified per genotype of rs10100651. Beanplots are shown for all phenotypes from the optimal multiple-phenotype association model with the largest effect size (C-allele as effect allele). Effect sizes and standard errors are shown above each plot. Samples were stratified on hard-called genotypes instead of genotype dosages as this SNP was imputed with high reliability (R^2 = 1.00) (x-axis), and intermediate phenotype residuals are shown for each individual (y-axis, small horizontal lines) as well as the mean per genotype (large horizontal line). Sample sizes per genotype are N = 94 (G/G), N = 146 (G/C) and N = 67 (C/C).



Supplementary Figure 5. eQTL effects and chromatin interactions for the discovered locus on chromosome 8. Circos plot showing association of variants in a segment of chromosome 8 (all variants tested with the optimal model for rs10100651, blue area highlights the LD block around rs10100651). Gene names are highlighted when rs10100651 exerts only eQTL effects (green), when an area including (SNPs in LD with) rs10100651 show only chromatin interaction (orange) or when both eQTL effects and chromatin interactions are present (red). (The genomic area including) rs10100651 shows both a significant eQTL effect and significant chromatin interaction for genes marked with * (also see Supplementary Tables 7 and 8). Plot generated with FUMA [25].



Supplementary Figure 6. Quantile-quantile plot of UK Biobank replication GWAS after SNP QC. A) Expected versus observed p-values for all SNPs in the UK Biobank PTSD GWAS after removal of SNPs with a minor allele count < 25 and a MAF < 0.05. λ_{GC} indicates inflation of observed p-values over expected p-values, whereas λ_{GC1000} indicates inflation for the dataset scaled to a GWAS with 1000 cases and 1000 controls. **B)** Regional association plot of the chromosome 8 locus in the UK biobank GWAS on self-reported PTSD case-control status. rs10100651 is in moderate LD with the top variant of this locus in the UK biobank (rs12550765, R^2 = 0.34 in the European population, R^2 = 0.47 in the British population).

3.5.2 Supplementary tables

Supplementary Table 1. Overview of surveys and questionnaires used in this study.

Questionnaire	Description
Revised Symptom Checklist (SCL-90-R) [15]	Contains 90 items rated on a 5-point scale ranging from 1 (not at all) to 5 (very much) and is a multidimensional measurement of psychiatric and somatic complaints, including the subdomains agoraphobia (7 items), anxiety (10 items), depression (16 items), somatic complaints (12 items), insufficiency in thinking and acting (9 items), sensitivity (18 items), hostility (6 items), sleeping problems (3 items), and eight miscellaneous items.
Self-Rating Inventory PTSD (SRIP) [16,17]	Contains 22 items that correspond to symptoms in the clusters re-experiencing, avoidance and numbing, and arousal as defined by the DSM-IV. Symptom severity (range 22-88) over the past month was assessed on a 4-point Likert scale ranging from 1 (not at all) to 4 (very much).
Checklist Individual Strength (CIS) [18]	Consists of 20 statements covering domains of subjective tiredness (8 items), concentration (5 items), motivation (4 items) and physical activity (3 items) as experienced in the previous 2 weeks, measured on a 7-point scale per statement where higher scores indicate more overall fatigue, concentration problems, reduced motivation and activity level.
Checklist potentially traumatic events [2]	Counts the number of items experienced during deployment, including: enemy fire, witnessed people suffering, witnessed wounded, colleague injured or killed, incoming fire, witnessed dead, rejected by locals, personal danger, witnessed others injured or killed, heard people screaming, insufficient means to intervene, mission experienced as useless, insufficient control over situation, memories of earlier deployments, traffic accident, held at gunpoint, physical injuries, colleague held hostage, held hostage.
Early Trauma Inventory (ETI) [19]	Consists of 27 dichotomous items that probe the severity of experienced trauma during childhood (until the age of 18).

Supplementary Table 2. Pre-imputation quality control of genetic data.

SNP-level quality control steps	**Number of SNPs**
SNPs pre-QC	713,014
Markers not coded with rs numbers	127
SNPs with heterozygous haploid genotypes	13,981
SNPs with genotyping rate < 0.95	4,339
SNPs failed HWE criteria ($p < 10^{-6}$)	2,178
Non-autosomal SNPs (excl. het. haploid)	5,524
Stand-ambiguous (A/T or C/G) SNPs	2,283
SNPs not in HRC imputation reference panel	6,584
Palindromic SNPs with MAF > 0.4	22
SNPs discordant with HRC (non-matching alleles)	2,096
SNPs MAF difference HRC > 0.15	427
Observed MAF vs. HRC (R^2)	0.984
SNPs post-QC	675,453

Individual-level quality control steps	**Number of individuals**
Individuals pre-QC	1,015 (926 male; 89 female)
Individuals genotyping rate < 0.95	10
Individuals discordant sex	3
Individuals excess heterozygosity (> 3 SD mean)	23
Individuals excess homozygosity (< 3 SD mean)	8
Individuals randomly removed due to relatedness (PIHAT > 0.10)	8
Individuals post-QC	963 (877 male; 86 female)
Overall genotyping rate pre-QC	0.995
Overall genotyping rate post-QC	0.999

Steps in *italic* were implemented in an HRC pre-imputation checking tool as recommended on the Michigan Imputation Server (www.well.ox.ac.uk/~wrayner/tools/#Checking). Steps are shown in the order in which they were performed, split on SNP-level QC and individual-level QC (individual level QC was performed before SNP-level QC).

Abbreviations: HWE, Hardy-Weinberg equilibrium; HRC, Haplotype Reference Consortium; MAF, minor-allele

Supplementary Table 3. Post-imputation quality control of genetic data.

Post-imputation quality control steps	Number of SNPs
SNPs pre-imputation	675,453
SNPs post-imputation	39,127,565
SNPs imputation $R^2 < 0.3$	18,048,362
SNPs MAF < 0.05	33,662,735
SNPs MAF difference HRC > 0.15	1,291,837
SNPs post-QC	5,361,288

Abbreviations: SNP: Single-nucleotide polymorphism; MAF, Minor allele frequency; HRC: Haplotype Reference Consortium

3

Supplementary Table 4. Model selection based on Bayesian Information Criterion for genome-wide significant hits.

Chr	Basepair	SNP	P-value	Model	BIC	BIC null
8	95887225	rs10100651	**9.9×10^{-9}**	**DHEA-S + GABA + NPY + SHBG + SCL-90-R + SRIP**	**94.44**	**108.45**
			1.9×10^{-8}	DHEA-S + GABA + NPY + SHBG + CIS + SCL-90-R + SRIP	95.12	104.56
			2.9×10^{-8}	DHEA-S + GABA + SHBG + SCL-90-R + SRIP	97.32	111.98
			3.1×10^{-8}	DHEA-S + GABA +NPY + SHBG + SRIP	108.79	122.98
14	47883746	rs4900759	**1.9×10^{-8}**	**SRIP + SCL-90-R**	**-15.22**	**8.45**
			3.6×10^{-8}	Cortisol + SCL-90-R + SRIP	-13.30	6.61

BIC estimates for models contributing to genome-wide association are shown. The preferred model is that with the lowest BIC score.

Supplementary Table 5. Effect sizes of phenotypes in optimal association models for the main analysis and sensitivity analyses.

Main analysis								
Chr	**Basepair**	**Effect allele**	**SNP**	**N**	**P-value**	**Phenotype in optimal model**	**Beta**	**SE**
8	95887225	C	rs10100651	307	9.9×10^{-9}	DHEA-S	1.2×10^{-1}	2.3×10^{-2}
						GABA	-6.5×10^{-3}	2.2×10^{-3}
						NPY	-4.7×10^{-3}	3.0×10^{-3}
						SHBG	-4.8×10^{-2}	1.2×10^{-2}
						SRIP	1.1×10^{-3}	1.2×10^{-2}
						SCL-90-R	-2.1×10^{-3}	5.0×10^{-3}
14	47883746	T	rs4900759	381	1.9×10^{-8}	SRIP	-5.5×10^{-2}	9.0×10^{-3}
						SCL-90-R	1.3×10^{-2}	3.9×10^{-3}

Log-transformed questionnaire data								
Chr	**Basepair**	**Effect allele**	**SNP**	**N**	**P-value**	**Phenotype in optimal model**	**Beta**	**SE**
8	95887225	C	rs10100651	312	2.9×10^{-8}	*DHEA-S*	1.2×10^{-1}	2.3×10^{-2}
						GABA	-6.2×10^{-3}	2.2×10^{-3}
						NPY	-4.1×10^{-3}	3.0×10^{-3}
						SHBG	-4.6×10^{-2}	1.2×10^{-2}
						SRIP	5.5×10^{-2}	3.2×10^{-1}
						SCL-90-R	8.7×10^{-2}	5.3×10^{-1}
14	47883746	T	rs4900759	337	1.5×10^{-5}	GABA	1.3×10^{-3}	1.8×10^{-3}
						NPY	1.9×10^{-3}	2.4×10^{-3}
						SRIP	1.4	4.3×10^{-1}
						SCL-90-R	-1.4	2.6×10^{-1}

< **Supplementary Table 5 continued.**

Questionnaire subscales								
Chr	**Basepair**	**Effect allele**	**SNP**	**N**	**P-value**	**Phenotype in optimal model**	**Beta**	**SE**
8	95887225	C	rs10100651	309	3.2×10^{-9}	*DHEA-S*	1.3×10^{-1}	2.3×10^{-2}
						GABA	-5.8×10^{-3}	2.2×10^{-3}
						NPY	-5.0×10^{-3}	2.9×10^{-3}
						SHBG	-4.6×10^{-2}	1.2×10^{-2}
						SRIP (avoidance)	-4.5×10^{-2}	2.3×10^{-2}
						SRIP (re-experiencing)	-3.8×10^{-3}	5.6×10^{-2}
						SCL-90-R (sleeping problems)	2.7×10^{-2}	3.2×10^{-2}
14	47883746	T	rs4900759	330	1.6×10^{-6}	GABA	1.6×10^{-3}	1.8×10^{-3}
						NPY	1.0×10^{-3}	2.4×10^{-3}
						SRIP (avoidance)	-5.2×10^{-2}	1.9×10^{-2}
						SRIP (re-experiencing)	-1.6×10^{-1}	4.1×10^{-2}
						SCL-90-R (sensitivity)	3.0×10^{-2}	1.3×10^{-2}

Effect sizes (betas) refer to the effect of the phenotype in the model with regard to the effect allele. For genome-wide significant hits in the main analysis (first panel), the analysis was repeated using log-transformed total scores for questionnaires (second panel) or subscales of questionnaires as a replacement for questionnaire total scores (third panel). Highlighting in *italic* indicates that the phenotype in the optimal model of the sensitivity analysis was also part of the optimal model in the main analysis.

Supplementary Table 6. Haplotype frequency analysis for rs10100651.

				Haplotypes				
SNP	**Genomic location relative to rs10100651 (bp)**	**I/G**	**R^2 with rs10100651**	**H1**	**H2**	**H3**	**H4**	**H5**
rs4075835	-53403	I	0.87	T	C	C	C	T
rs7817862	-39351	I	0.87	C	G	G	G	C
rs13271644	-33277	I	0.86	A	G	G	G	G
rs10441538	-18617	I	0.87	G	A	A	A	G
rs7002058	-13232	I	0.87	G	C	C	C	G
rs713113	-9438	I	0.86	T	C	C	C	C
rs10100651	0	I	1.0	G	C	C	G	G
rs1129152	1082	G	0.76	C	T	C	C	C
Haplotype frequency in cohort				0.508	0.391	0.065	0.034	0.002
Haplotype frequency in 1000 genomes				0.454	0.457	0.034	0.044	0

The first two columns show the SNP rs-numbers and their relative position on chromosome 8 compared to rs10100651. The third column shows whether the SNP was imputed (I) or directly genotyped on the array (G). The fourth column shows linkage disequilibrium with rs10100651 (R^2) in our dataset. The following five columns show different haplotypes (H1-H5) with their respective frequencies in the sample with which the association was found and in the European subset of 1000 genomes phase 3 (bottom rows).

Supplementary Table 7. Significant eQTL effects of rs10100651.

Database	Tissue	P	Signed stat.	FDR	Gene
BIOSQTL	BIOS_eQTL_geneLevel	5.07E-19	-8.91	0	*TP53INP1*
BIOSQTL	BIOS_eQTL_geneLevel	2.39E-45	14.13	0	*INTS8*
BIOSQTL	BIOS_eQTL_geneLevel	4.19E-11	6.60	0	*INTS8*
BIOSQTL	BIOS_eQTL_geneLevel	4.19E-11	6.60	0	*CCNE2*
GTEx_v7	Thyroid	3.04E-09	0.20	2.04E-22	*INTS8*
GTEx_v7	Lung	4.83E-12	0.19	6.32E-12	*INTS8*
GTEx_v7	Adipose_Subcutaneous	1.71E-05	0.15	7.55E-12	*INTS8*
GTEx_v7	Skin_Sun_Exposed_Lower_leg	2.32E-11	0.17	2.35E-11	*TP53INP1*
GTEx_v7	Lung	7.93E-09	0.32	7.56E-10	*CCNE2*
GTEx_v7	Pancreas	4.83E-09	0.33	3.49E-09	*INTS8*
GTEx_v7	Heart_Left_Ventricle	2.44E-06	0.16	1.14E-07	*NDUFAF6*
GTEx_v7	Artery_Tibial	5.74E-10	0.17	4.61E-07	*INTS8*
GTEx_v7	Skin_Not_Sun_Exposed_Suprapubic	2.20E-07	0.18	1.62E-06	*TP53INP1*
GTEx_v7	Artery_Aorta	2.34E-05	0.17	2.62E-06	*INTS8*
GTEx_v7	Esophagus_Muscularis	1.22E-05	0.12	6.07E-06	*INTS8*
GTEx_v6	Thyroid	5.19E-06	4.66	2.24E-04	*INTS8*
GTEx_v7	Testis	5.78E-08	-0.20	2.50E-04	*CCNE2*
GTEx_v6	Skin_Sun_Exposed_Lower_leg	2.18E-09	6.20	2.88E-04	*TP53INP1*
GTEx_v6	Heart_Left_Ventricle	6.07E-06	4.69	5.97E-04	*NDUFAF6*
GTEx_v6	Pancreas	8.72E-07	5.21	6.13E-04	*INTS8*
GTEx_v6	Lung	4.53E-07	5.19	7.12E-04	*INTS8*
GTEx_v6	Heart_Atrial_Appendage	6.49E-06	4.71	7.21E-04	*CCNE2*
GTEx_v6	Artery_Tibial	3.65E-06	4.74	9.17E-04	*INTS8*
GTEx_v6	Testis	9.21E-06	-4.63	9.94E-04	*CCNE2*
CMC	CMC_SVA_cis	NA	1.00	9.00E-03	*INTS8*

Results obtained from analysis in FUMA [25].
eQTL effects of rs10100651 (C-allele as effect allele) with FDR < 0.05 are shown.
BIOS_eQTL_geneLevel eQTL data was derived from whole blood, CMC_SVA_cis eQTL data was derived from dorsolateral prefrontal cortex.
Abbreviations: *INTS8*, Integrator Complex Subunit 8; *TP53INP1*, Tumor Protein P53 Inducible Nuclear Protein 1; *CCNE2*, Cyclin E2; *NDUFAF6*, NADH:Ubiquinone Oxidoreductase Complex Assembly Factor 6. BIOS, Biobank-Based Integrative Omics Study; GTEx, Genotype-Tissue Expression consortium; CMC, CommonMind Consortium; SVA, surrogate variable analysis

Supplementary Table 8. Significant chromatin interactions of rs10100651 locus.

Region 1	Region 2	FDR	Tissue/Cell	Genes
8:95880001-95920000	8:95840001-95880000	1.33E-105	Mesenchymal stem cell	NA
8:95880001-95920000	8:95840001-95880000	1.08E-92	hESC	NA
8:95880001-95920000	8:95840001-95880000	1.66E-66	IMR90	NA
8:95880001-95920000	8:95840001-95880000	3.41E-66	Mesendoderm	NA
8:95880001-95920000	8:96200001-96240000	5.87E-36	IMR90	NA
8:95880001-95920000	8:95640001-95680000	1.84E-34	hESC	*ESRP1*
8:95880001-95920000	8:96240001-96280000	5.12E-25	IMR90	NA
8:95880001-95920000	8:95680001-95720000	1.59E-23	hESC	NA
8:95880001-95920000	8:96200001-96240000	3.67E-17	Mesenchymal stem cell	NA
8:95880001-95920000	8:95840001-95880000	2.37E-16	GM12878	NA
8:95880001-95920000	8:96240001-96280000	4.87E-15	Mesenchymal stem cell	NA
8:95880001-95920000	8:95840001-95880000	4.91E-15	Left_Ventricle	NA
8:95880001-95920000	8:96040001-96080000	7.94E-14	IMR90	NA
8:95880001-95920000	8:96200001-96240000	9.18E-14	GM12878	NA
8:95880001-95920000	8:96000001-96040000	3.03E-13	GM12878	NA
8:95880001-95920000	8:96000001-96040000	3.48E-13	Mesenchymal stem cell	NA
8:95880001-95920000	8:96240001-96280000	9.07E-13	GM12878	NA
8:95880001-95920000	8:96120001-96160000	3.60E-12	GM12878	*PLEKHF2*
8:95880001-95920000	8:96160001-96200000	1.91E-10	GM12878	NA
8:95880001-95920000	8:96280001-96320000	2.12E-10	IMR90	*C8ORF37*
8:95880001-95920000	8:96280001-96320000	3.28E-10	Left_Ventricle	*C8ORF37*
8:95880001-95920000	8:95840001-95880000	4.64E-10	Liver	NA
8:95880001-95920000	8:96200001-96240000	4.97E-10	Mesendoderm	NA
8:95880001-95920000	8:95640001-95680000	6.97E-10	Mesendoderm	*ESRP1*
8:95880001-95920000	8:96160001-96200000	1.49E-09	Mesenchymal stem cell	NA
8:95880001-95920000	8:95800001-95840000	5.28E-09	IMR90	*INTS8*
8:95880001-95920000	8:96280001-96320000	7.57E-09	GM12878	*C8ORF37*
8:95880001-95920000	8:95840001-95880000	9.36E-09	Trophoblast-like cell	NA
8:95880001-95920000	8:97200001-97240000	2.12E-08	Mesenchymal stem cell	NA
8:95880001-95920000	8:96200001-96240000	5.77E-08	Trophoblast-like cell	NA
8:95880001-95920000	8:97200001-97240000	5.81E-08	IMR90	NA
8:95880001-95920000	8:95640001-95680000	1.72E-07	Trophoblast-like cell	*ESRP1*

< Supplementary Table 8 continued.

Region 1	Region 2	FDR	Tissue/Cell	Genes
8:95880001-95920000	8:96160001-96200000	1.77E-07	IMR90	NA
8:95880001-95920000	8:96120001-96160000	2.15E-07	IMR90	*PLEKHF2*
8:95880001-95920000	8:96200001-96240000	2.41E-07	hESC	NA
8:95880001-95920000	8:97280001-97320000	5.19E-07	IMR90	NA
8:95880001-95920000	8:95960001-96000000	9.87E-07	GM12878	*TP53INP1*
8:95880001-95920000	8:95800001-95840000	9.87E-07	GM12878	*INTS8*

Results obtained from analysis in FUMA [25].
Interactions from GSE87112 database. Region 1 is the region containing rs10100651 and surrounding variants in linkage disequilibrium. Genes column lists genes which promoter regions overlap with region 2.
In the tissue/cell column, GM12878 cells are B-lymphocytes from blood, IMR90 cells are derived from lung tissue, and hESC cells are human embryonic stem cells.

Supplementary Table 9. Significant associations from *TP53INP1* phenome-wide analysis.

atlas ID	Study ID	Year	Domain	Trait	P-value	N
3328	UKB	2017	Endocrine	Diabetes (diagnosed by doctor)	3.98E-09	385,420
3449	UKB	2017	Metabolic	Impedance measures - Impedance of leg (left)	1.47E-08	379,807
3448	UKB	2017	Metabolic	Impedance measures - Impedance of leg (right)	3.34E-08	379,813
3786	BioRxiv: doi.org/10.1101/214973	2017	Neurological	Insomnia	6.65E-08	386,533
3597	UKB	2017	Cardiovascular	Non-cancer illness code, self-reported: hypertension	1.26E-07	289,307
3601	UKB	2017	Endocrine	Non-cancer illness code, self-reported: diabetes	3.54E-07	289,307
3745	UKB	2017	Psychiatric	Happiness and subjective well-being - General happiness	7.82E-07	126,132
3232	UKB	2017	Neurological	Sleeplessness / insomnia	3.26E-06	386,078
3447	UKB	2017	Metabolic	Impedance measures - Impedance of whole body	7.14E-06	379,792
3787	BioRxiv: doi.org/10.1101/214973	2017	Neurological	Insomnia (male)	8.05E-06	208,716
3551	UKB	2017	Cardiovascular	Vascular/heart problems diagnosed by doctor: High blood pressure	8.59E-06	385,699

Abbreviations: atlas ID, identifier of the GWAS in GWAS atlas (atlas.ctglab.nl); PMID, Study ID in PubMed, or link to BioRxiv. UKB indicates unpublished GWAS based on UK Biobank trait included in GWAS atlas (atlas.ctglab.nl); Year, year the GWAS (or data on which the GWAS is based) was published; Domain, phenotypic domain of the trait; Trait, description of the phenotype tested in the GWAS; P-value, association p-value of *TP53INP1* with the tested trait; N, GWAS sample size.

Supplementary Table 10. Significant associations from *INTS8* phenome-wide analysis.

atlas ID	Study ID	Year	Domain	Trait	P-value	N
3379	UKB	2017	Cardiovascular	Diastolic blood pressure (automated reading)	3.25E-08	361,411
3597	UKB	2017	Cardiovascular	Non-cancer illness code, self-reported: hypertension	3.42E-07	289,307
3186	UKB	2017	Metabolic	Hip circumference	4.06E-07	385,887
3453	UKB	2017	Metabolic	Impedance measures - Leg fat mass (right)	5.24E-07	379,802
3445	UKB	2017	Metabolic	Impedance measures - Body mass index (BMI)	5.69E-07	379,831
3440	UKB	2017	Metabolic	Impedance measures - Weight	9.29E-07	379,840
3457	UKB	2017	Metabolic	Impedance measures - Leg fat mass (left)	1.17E-06	379,783
3435	UKB	2017	Metabolic	Body mass index	1.76E-06	385,336
3436	UKB	2017	Metabolic	Weight	2.50E-06	385,473
3461	UKB	2017	Metabolic	Impedance measures - Arm fat mass (right)	2.53E-06	379,725
3452	UKB	2017	Metabolic	Impedance measures - Leg fat percentage (right)	3.11E-06	379,806
3465	UKB	2017	Metabolic	Impedance measures - Arm fat mass (left)	3.76E-06	379,663
3442	UKB	2017	Metabolic	Impedance measures - Whole body fat mass	4.64E-06	379,203
3551	UKB	2017	Cardiovascular	Vascular/heart problems diagnosed by doctor: High blood pressure	5.63E-06	385,699
3467	UKB	2017	Metabolic	Impedance measures - Arm predicted mass (left)	6.22E-06	379,638

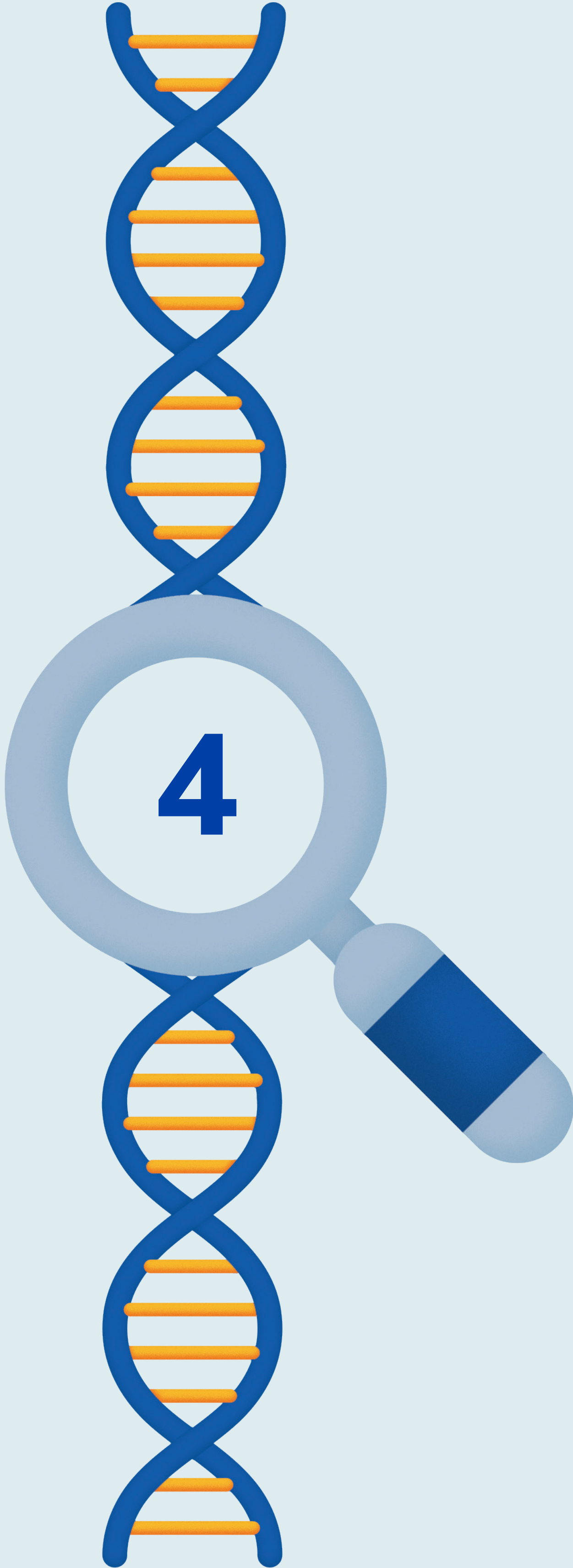
Abbreviations: atlas ID, identifier of the GWAS in GWAS atlas (atlas.ctglab.nl); PMID, Study ID in PubMed, or link to BioRxiv. UKB indicates unpublished GWAS based on UK Biobank trait included in GWAS atlas (atlas.ctglab.nl); Year, year the GWAS (or data on which the GWAS is based) was published; Domain, phenotypic domain of the trait; Trait, description of the phenotype tested in the GWAS; P-value, association p-value of *INTS8* with the tested trait; N, GWAS sample size.

3.6 References

1. Lapierre CB, Schwegler AF, Labauve BJ. Posttraumatic stress and depression symptoms in soldiers returning from combat operations in Iraq and Afghanistan. J Trauma Stress 2007;20(6):933–43.
2. Reijnen A, Rademaker AR, Vermetten E, Geuze E. Prevalence of mental health symptoms in Dutch military personnel returning from deployment to Afghanistan: a 2-year longitudinal analysis. Eur Psychiatry 2015;30(2):341–6.
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4 ed. Washington, DC: 2000.
4. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5 ed. Washington, D.C.: 2013.
5. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, et al. DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. Am J Psychiatry 2013;170(1):59–70.
6. Flint J, Timpson N, Munafò M. Assessing the utility of intermediate phenotypes for genetic mapping of psychiatric disease. Trends Neurosci 2014;37(12):733–41.
7. Magi R, Suleimanov YV, Clarke GM, Kaakinen M, Fischer K, Prokopenko I, et al. SCOPA and META-SCOPA: software for the analysis and aggregation of genome-wide association studies of multiple correlated phenotypes. BMC Bioinformatics 2017;18(1):25.
8. O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin M-R, et al. MultiPhen: Joint Model of Multiple Phenotypes Can Increase Discovery in GWAS. PLoS One 2012;7(5):e34861.
9. World Medical Association. World Medical Association Declaration of Helsinki. JAMA 2013;310(20):2191–4.
10. van Zuiden M, Haverkort SQ, Tan Z, Daams J, Lok A, Olff M. DHEA and DHEA-S levels in posttraumatic stress disorder: A meta-analytic review. Psychoneuroendocrinology 2017;84:76–82.
11. Geuze E, van Wingen GA, van Zuiden M, Rademaker AR, Vermetten E, Kavelaars A, et al. Glucocorticoid receptor number predicts increase in amygdala activity after severe stress. Psychoneuroendocrinology 2012;37(11):1837–44.
12. Schür RR, Boks MP, Geuze E, Prinsen HC, Verhoeven-Duif NM, Joëls M, et al. Development of psychopathology in deployed armed forces in relation to plasma GABA levels. Psychoneuroendocrinology 2016;73:263–70.
13. Reijnen A, Geuze E, Eekhout I, Maihofer AX, Nievergelt CM, Baker DG, et al. Biological profiling of plasma neuropeptide Y in relation to posttraumatic stress symptoms in two combat cohorts. Biol Psychol 2018;134:72–9.
14. Reijnen A, Geuze E, Vermetten E. The effect of deployment to a combat zone on testosterone levels and the association with the development of posttraumatic stress symptoms: A longitudinal prospective Dutch military cohort study. Psychoneuroendocrinology 2015;51:525–33.
15. Arrindell WA, Ettema J. Revised manual for a multidimensional indicator of psychopathology In: Herziene handleiding bij een multidimensionele psychopathologie indicator. Swets Zeitlinger, Lisse, the Netherlands 2003;
16. Hovens JE, Bramsen I, van der Ploeg HM. Self-rating inventory for posttraumatic stress disorder: review of the psychometric properties of a new brief Dutch screening instrument. Percept Mot Skills 2002;94(3 Pt 1):996–1008.
17. Hovens JE, Bramsen I, van der Ploeg HM. Self-report measure for PTSD symptoms: SRIP manual In: De Zelfinventarisatielijst Post traumatische Stressstoornis: ZIL handleiding. Swets Zeitlinger, Lisse, the Netherlands 2000;
18. Vercoulen J, Alberts M, Bleijenberg G. De checklist individuele spankracht (CIS). Gedragstherapie 1999;32:131–6.
19. Bremner JD, Bolus R, Mayer EA. Psychometric properties of the Early Trauma Inventory-Self Report. J Nerv Ment Dis 2007;195(3):211–8.

20. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.
21. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48(10):1284–7.
22. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48(10):1279–83.
23. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 2016;48(11):1443–8.
24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet 2011;88(1):76–82.
25. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8(1):1826.
26. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015;348(6235):648–60.
27. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 2017;550(7675):204–13.
28. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013;45(10):1238–43.
29. Zhernakova DV, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet 2017;49(1):139–45.
30. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 2014;17(10):1418–28.
31. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 2012;44(10):1084–9.
32. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat Neurosci 2017;20(10):1418–26.
33. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat Neurosci 2016;19(11):1442–53.
34. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep 2016;17(8):2042–59.
35. Watanabe K, Stringer S, Frei O, Umićević Mirkov M, Polderman TJC, van der Sluis S, et al. A global view of pleiotropy and genetic architecture in complex traits. bioRxiv 2018;
36. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. Mol Psychiatry 2018;23(3):666–73.
37. Neale Lab. http://www.nealelab.is/uk-biobank/ [Internet]. [cited 2018 Aug 20];Available from: http://www.nealelab.is/uk-biobank/
38. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779.
39. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21(2):263–5.
40. The 1000 Genomes Project Consortium C. A global reference for human genetic variation. Nature 2015;526(7571):68–74.
41. Yehuda R, Brand SR, Golier JA, Yang RK. Clinical correlates of DHEA associated with post-traumatic stress disorder. Acta Psychiatr Scand 2006;114(3):187–93.
42. Russo SJ, Murrough JW, Han M-H, Charney DS, Nestler EJ. Neurobiology of resilience. Nat Neurosci 2012;15(11):1475–84.

43. Kimonides VG, Khatibi NH, Svendsen CN, Sofroniew MV, Herbert J. Dehydroepiandrosterone (DHEA) and DHEA-sulfate (DHEAS) protect hippocampal neurons against excitatory amino acid-induced neurotoxicity. Proc Natl Acad Sci U S A 1998;95(4):1852–7.
44. Bastianetto S, Ramassamy C, Poirier J, Quirion R. Dehydroepiandrosterone (DHEA) protects hippocampal cells from oxidative stress-induced damage. Brain Res Mol Brain Res 1999;66(1-2):35–41.
45. Vaiva G, Boss V, Ducrocq F, Fontaine M, Devos P, Brunet A, et al. Relationship between posttrauma GABA plasma levels and PTSD at 1-year follow-up. Am J Psychiatry 2006;163(8):1446–8.
46. Pitman RK, Rasmusson AM, Koenen KC, Shin LM, Orr SP, Gilbertson MW, et al. Biological studies of post-traumatic stress disorder. Nat Rev Neurosci 2012;13(11):769–87.
47. Oegema R, Baillat D, Schot R, van Unen LM, Brooks A, Kia SK, et al. Human mutations in integrator complex subunits link transcriptome integrity to brain development. PLoS Genet 2017;13(5):e1006809.
48. Saadi H, Seillier M, Carrier A. The stress protein TP53INP1 plays a tumor suppressive role by regulating metabolic homeostasis. Biochimie 2015;118:44–50.
49. Miller MW, Lin AP, Wolf EJ, Miller DR. Oxidative Stress, Inflammation, and Neuroprogression in Chronic PTSD. Harv Rev Psychiatry 2018;26(2):57–69.
50. Edmondson D, Känel von R. Post-traumatic stress disorder and cardiovascular disease. Lancet Psychiatry 2017;4(4):320–9.
51. Mellman TA, Pigeon WR, Nowell PD, Nolan B. Relationships between REM sleep findings and PTSD symptoms during the early aftermath of trauma. J Trauma Stress 2007;20(5):893–901.
52. Berle D, Hilbrink D, Russell-Williams C, Kiely R, Hardaker L, Garwood N, et al. Personal wellbeing in posttraumatic stress disorder (PTSD): association with PTSD symptoms during and following treatment. BMC Psychol 2018;6(1):594.
53. Almli LM, Stevens JS, Smith AK, Kilaru V, Meng Q, Flory J, et al. A genome-wide identified risk variant for PTSD is a methylation quantitative trait locus and confers decreased cortical activation to fearful faces. Am J Med Genet B Neuropsychiatr Genet 2015;168B(5):327–36.
54. Morey RA, Davis SL, Garrett ME, Haswell CC, Mid-Atlantic MIRECC Workgroup, Marx CE, et al. Genome-wide association study of subcortical brain volume in PTSD cases and trauma-exposed controls. Transl Psychiatry 2017;7(11):1265.
55. Schmidt U, Willmund G-D, Holsboer F, Wotjak CT, Gallinat J, Kowalski JT, et al. Searching for non-genetic molecular and imaging PTSD risk and resilience markers: Systematic review of literature and design of the German Armed Forces PTSD biomarker study. Psychoneuroendocrinology 2015;51:444–58.
56. Nievergelt CM, Ashley-Koch AE, Dalvie S, Hauser MA, Morey RA, Smith AK, et al. Genomic Approaches to Posttraumatic Stress Disorder: The Psychiatric Genomic Consortium Initiative. Biol Psychiatry 2018;83(10):831–9.
57. Liberzon I. Searching for Intermediate Phenotypes in Posttraumatic Stress Disorder. Biol Psychiatry 2018;83(10):797–9.
58. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature 2010;467(7311):52–8.


4

Chapter 4

The effect of genetic vulnerability and military deployment on the development of posttraumatic stress disorder and depressive symptoms

Dick Schijven*, Remmelt R. Schür*, Marco P. Boks, Bart P.F. Rutten, Murray B. Stein,
Jan H. Veldink, Marian Joëls, Elbert Geuze, Eric Vermetten, Jurjen J. Luykx, Christiaan H. Vinkers

*Equal contributions

Manuscript submitted

Abstract

Exposure to trauma strongly increases the risk to develop stress-related psychopathology, such as posttraumatic stress disorder (PTSD) or major depressive disorder (MDD). In addition, liability to develop these moderately heritable disorders is partly determined by common genetic variance, which is starting to be uncovered by genome-wide association studies (GWASs). However, it is currently unknown to what extent genetic vulnerability and trauma interact. We investigated whether genetic risk based on summary statistics of large GWASs for PTSD and MDD predisposed individuals to report an increase in MDD and PTSD symptoms in a prospective military cohort (N = 516) at five time points after deployment to Afghanistan: one month, six months and one, two and five years. Linear regression was used to analyze the contribution of polygenic risk scores (PRSs, at multiple p-value thresholds of association) and their interaction with deployment-related trauma to the development of PTSD- and depression-related symptoms. We found no main effects of PRSs nor evidence for interaction with trauma on the development of PTSD or depressive symptoms at any of the available time points in the five years after military deployment. Our results based on a unique long-term follow-up of a deployed military cohort suggest limited validity of current PTSD and MDD polygenic risk scores, albeit in the presence of minimal severe psychopathology in the target cohort. Even though the predictive value of PRSs will likely benefit from larger sample sizes in discovery and target datasets, progress will probably also depend on (endo) phenotype refinement that in turn will reduce etiological heterogeneity.

4.1 Introduction

Deployment to a combat zone constitutes a major risk factor for the development of several debilitating psychiatric disorders such as posttraumatic stress disorder (PTSD) and major depressive disorder (MDD) [1,2]. Nevertheless, only a relatively small proportion of individuals develop psychopathology after traumatic experiences during deployment and the majority is resilient and remains unaffected, depending on factors such as trauma severity or the number of experienced events [3-5]. These different outcomes can partly be explained by common genetic variation. Compelling evidence suggests that both PTSD and MDD have a heritable component, with single-nucleotide polymorphism (SNP)-based estimates ranging from 0.07 (in males) - 0.29 (in females) [6] and 0.09 [7], respectively. However, it is currently unknown to what extent genetic risk interacts with traumatic experiences in increasing the risk for MDD or PTSD.

Recently, large genome-wide association studies (GWASs) were published for both PTSD [6,8] and MDD [7,9], aiming to elucidate the role of common genetic variation in these disorders. Results from these studies provide estimates of genetic risk for these disorders and can be used to generate polygenic risk scores (PRSs) in other, independent, cohorts [10]. The use of PRSs for PTSD and MDD in a large longitudinal cohort exposed to trauma may further elucidate the interaction of genetic vulnerability and trauma exposure on the development of psychopathology and could be of value to predict vulnerability.

Here, we hypothesized that genetic susceptibility for increase in self-reported PTSD- and MDD-related symptoms (as indicated by higher PRSs) in interaction with trauma exposure during military deployment determines the degree to which these symptoms become manifest in a period of up to five years after deployment. In addition, we explored relations of PRSs with symptom levels prior to deployment and the interaction of PRSs with childhood trauma on pre- and the development of post-deployment symptom levels.

4.2 Materials and methods

4.2.1 Participants and general procedure

Participants of the present study were military deployed to Afghanistan for 4 months between 2005 and 2008, either as part of a Provincial Reconstruction Team or as part of the International Security Assistance Force (see outlined in Reijnen *et al.* and van Zuiden *et al.* [2,11]). All 1,032 participants gave both written and oral informed consent. Assessments and data collection took place prior to, as well as one month, six months and one, two and five years after deployment. A previous study provides general information about common duties and deployment-related potentially traumatic events [2]. Study approval was granted by the University Medical Center Utrecht (UMCU) Institutional Review Board.

4.2.2 Genetic data

DNA isolated from blood of 1,015 participants was genotyped on the Illumina Human OmniExpress-24 v1.1 BeadChip array (Illumina, San Diego, CA, USA). A total number of 713,040 markers was present on this chip. Data underwent extensive pre-imputation quality control (QC, see Supplementary Table 1). First, we removed subjects with a genotype call rate < 0.95, discordant genetic and reported sex in phenotype data, excess heterozygosity rate (heterozygosity > 3SD from the mean) or excess homozygosity (heterozygosity < 3SD from the mean) and a relatedness coefficient (PIHAT) > 0.1 (one of a pair of related individuals was removed at random). Second, we removed non-autosomal variants, SNPs with a genotype call rate < 0.95, SNPs out of Hardy-Weinberg Equilibrium (HWE test p-value < 1×10^{-6}) and SNPs with a strand-ambiguous AT/CG genotype. Finally, as a preparatory step for imputation, we removed SNPs that were not present in the imputation reference panel or had a difference in minor allele frequency (MAF) > 0.15 compared to the reference panel. A total number of 963 individuals and 675,453 SNPs remained after pre-imputation QC.

We imputed the data on the Michigan Imputation Server [12] using the Haplotype Reference Consortium Release 1.1 [13] as a reference panel and Eagle version 2.3 as a phasing algorithm [14]. We performed post-imputation QC, where we removed SNPs that had a MAF < 0.05, imputation R^2 < 0.3 and MAF difference

from the reference panel exceeding 0.15 (see Supplementary Table 2). After post-imputation QC, the data contained a total of 5,361,288 SNPs. We extracted genotype dosages from the imputed data for individuals with a Central European (CEU) ancestry according to the first two principal components of the genetic data using HapMap3 populations as a reference (see Supplementary Figure 1). We only included individuals with European ancestry (constituting the vast majority of our cohort), to match the ancestry of discovery datasets used for PRS calculation and thus rule out confounding due to population stratification in our analyses.

4.2.3 Discovery GWAS data

We obtained summary statistics for recent GWASs of PTSD and MDD. For PTSD, this included results for three cohorts within the Army Study to Assess Risk and Resilience in Servicemembers (STARRS) (totaling 2,812 cases and 6,244 controls of European American Ancestry) [8] and results of the largest published PTSD GWAS (2,424 cases and 7,113 controls of European ancestry) [6]. For depression, we used data from the largest MDD case-control GWAS (135,458 cases and 344,901 controls) [7] and a large self-report depression GWAS (75,607 cases and 231,747 controls) [9]. Although the full results from Wray *et al.* included the Hyde *et al.* cohort, the publicly available summary statistics did not include these data due to sharing limitations.

From discovery GWAS datasets, we removed all variants other than SNPs (e.g. indels), non-autosomal SNPs, SNPs with MAF < 0.05, SNPs with imputation INFO score < 0.8, strand-ambiguous A/T or C/G genotype, and duplicate or multiallelic SNPs. For the Duncan *et al.* PTSD GWAS, allele frequencies and imputation INFO scores were not available in summary statistics. However, SNPs with INFO scores < 0.9 were reported to be removed during QC [6] and we filtered for SNPs with MAF ≥ 0.05 using frequencies calculated on the European subset of 1000 genomes phase 3 as a reference [15].

To improve the predictive power of our PRSs, we meta-analyzed both PTSD studies and both MDD studies. First, to exclude the possibility of confounding due to sample overlap between these studies, we used linkage disequilibrium score regression (LDSC) to calculate the bivariate intercept. In the absence of sample overlap the intercept is 0, but this value increases when shared samples

are present between GWAS datasets [16]. In LDSC, LD scores were calculated in 1 centiMorgan (cM) windows around HapMap3 SNPs using 1000 genomes phase 3 as a reference for LD [15]. Second, we applied inverse variance-weighted fixed effects meta-analysis in METAL [17], generated quantile-quantile plots of p-value distributions for PTSD and MDD meta-analyses and calculated the genomic inflation factor (lambda). For MDD [7], association statistics for a limited set of ten thousand independent top SNPs including Hyde *et al.* data were publicly available. We used these data to calculate the correlation between effect sizes and association p-values of the original publication with those of our MDD meta-analysis (which is based on the same data). We used the meta-analyzed PTSD and MDD GWASs as discovery data for PRS calculation.

4.2.4 Polygenic risk score calculation

We ruled out the possibility of sample overlap between our target cohort and discovery GWAS datasets. Prior to PRS calculation, we removed linkage disequilibrium (LD) from the discovery data using a clumping procedure in PLINK version 1.90b3z [18] where all correlated SNPs in a genetic window, except the SNP with the lowest association p-value, were removed (first using a genetic window of 250 kilobase and an LD $R^2 > 0.5$ followed by a second run using a genetic window of 5000 kb and an LD $R^2 > 0.2$). Additionally, we excluded all SNPs in genomic regions with strong or complex LD structures (e.g. the MHC region on chromosome 6) (Supplementary Table 3). We used the European subset of 1000 genomes phase 3 as a reference for LD [15]. PTSD and MDD PRSs were calculated using the score function in PLINK for each of 13 sets of SNPs selected on different association p-value thresholds in the discovery GWAS (P_T): $< 5 \times 10^{-8}$, $< 5 \times 10^{-7}$, $< 5 \times 10^{-6}$, $< 5 \times 10^{-5}$, $< 5 \times 10^{-4}$, $< 5 \times 10^{-3}$, < 0.05, < 0.1, < 0.2, < 0.3, < 0.4, < 0.5 and ≤ 1). For PTSD, we were only able to calculate PRS for the latter eleven P_Ts, as an insufficient number of SNPs was available for the most stringent P_Ts. All scores were scaled around a mean of zero with a variance of one to allow for comparison of effect sizes between P_Ts.

4.2.5 Psychopathology outcomes

PTSD symptoms were measured prior to and one month, six months and one, two and five years after deployment, whereas for depressive symptoms all but

the five year post-deployment measurement were available. PTSD and depressive symptoms were evaluated using the Self-Report Inventory for PTSD (SRIP) [19] and the depression subscale of the symptom checklist-90 (SCL-90) [20], respectively. The correlation of change in depressive symptoms and PTSD symptoms was significant from pre- to post-deployment (one month: Pearson's rho (r) = 0.11, p = 0.03; six months: r = 0.41, $p = 1.3 \times 10^{-15}$; one year: r = 0.62, $p < 2.2 \times 10^{-16}$; two years: r = 0.56, $p < 2.2 \times 10^{-16}$). A self-report checklist described in a previous study [2] was used to assess exposure to potentially traumatic events, including direct combat. The self-report Early Trauma Inventory was used to assess potentially traumatic events during childhood (subdivided in physical, sexual and emotional abuse) [21].

4.2.6 Statistical analyses

Paired two-tailed t-tests were used to compare pre- and post-deployment psychopathology symptom levels. We first analyzed the association between PRSs and childhood- and deployment trauma (e.g. trauma during deployment ~ PRS PTSD P_T 0.05) using a multiple-testing corrected significance threshold for two measurements and two models tested ($p < 0.05/4 = 0.0125$).

For primary analyses, linear regression models were used with post-deployment PTSD symptom levels as the dependent variable and as indicators: pre-deployment PTSD symptom levels, a PRS for PTSD (11 different P_T, see above), trauma during deployment, childhood trauma, sex, age, and the five first principal components (adjusting for ethnicity) (e.g. PTSD symptoms post_six_months ~ PTSD symptoms pre + PRS PTSD P_T 0.05 + trauma during deployment + childhood trauma + sex + age + PC1 + PC2 + PC3 + PC4 + PC5). Subsequently, the interaction of the PTSD PRSs with trauma exposure during deployment was added to the model, in a model including all PRS x covariate and trauma exposure x covariate interactions to correct for possible confounding [22]. Identical models were used for depressive symptoms and MDD PRSs (13 P_T).

In secondary analyses, we examined the association of PRSs with pre-deployment psychopathology levels and the interaction of PRSs and childhood trauma on pre- and the development of post-deployment psychopathology levels. Homogeneity of variance, absence of outliers (Cook's Distance < 1) and normal distribution of the residuals were confirmed by inspecting plots of the

main models. To correct for multiple testing in primary and secondary analyses, the significance threshold was set at 0.05/18 = 0.0028 to adjust for analyses of five time points for PTSD symptoms and four for depressive symptoms using two models (one with only main effects and one including interaction).

4.3 Results

4.3.1 Descriptive statistics

Table 1 lists sample characteristics, including age, sex, number of traumatic events during military deployment and childhood, number of previous deployments, and pre- and post-deployment levels of PTSD and depressive symptoms. PTSD symptoms were significantly higher one ($p < 0.001$) and six ($p = 0.03$) months and five years ($p = 0.002$) after deployment compared to pre-deployment, but not one ($p = 0.95$) and two ($p = 0.21$) years after deployment. In contrast, depressive symptoms were significantly higher one ($p = 0.008$) and two years ($p = 0.0001$) post-deployment, whereas one month ($p = 0.13$) and six months ($p = 0.07$) post-deployment no significant increases were present compared to pre-deployment depressive symptoms (see Table 1).

4.3.2 Meta-analyses of discovery datasets and PRS calculation

Bivariate LDSC of PTSD summary statistics and MDD GWAS results did not reveal any intercepts more than one standard error above 0, indicating no suspected presence of sample overlap. Genetic correlation calculated with LDSC between MDD datasets was high ($r_g = 0.86$, se = 0.03) which is as expected for GWASs of the same phenotype (see Supplementary Table 4). Reliable genetic correlation estimates between PTSD datasets could not be calculated due to low sample size. Inflation of test statistics was absent in the meta-analysis of PTSD GWAS results (5,236 cases; 13,357 controls, $\lambda_{GC} = 1.011$; $\lambda_{GC1000} = 1.001$) while high inflation was observed in the MDD meta-analysis (135,458 cases; 344,901 controls; $\lambda_{GC} = 1.44$; $\lambda_{GC1000} = 1.002$) as reported in the original manuscript [7]. We found high correlations between effect estimates ($R^2 = 0.997$) and association p-values ($R^2 = 0.97$) of our meta-analysis and the original result for 8,398 SNPs for which full results were publicly available. PRSs for PTSD and MDD were calculated using meta-analysis results (PRS distributions per P_T are shown in Supplementary Figure 3 and Supplementary Figure 4).

Table 1. Sample characteristics (total N = 516).

Age (mean, SD)	29.5 (9.3)				
Sex (% male)	91.5				
Number of deployment-related trauma events (mean, SD)	4.6 (3.2)				
Number of early trauma events (mean, SD)	3.3 (3.0)				
Number of previous deployments (mean, SD)	0.9 (1.2)				
	Pre-deployment	**One month post-deployment**	**t**	**p**	**n**
PTSD symptoms (SRIP) (mean, SD)	26.8 (4.9)	27.7 (5.9)	-3.71	< 0.001	421
Depressive symptoms (SCL-90 depression subscale) (mean, SD)	17.9 (3.0)	18.1 (3.7)	-1.50	0.13	486
	Pre-deployment	**Six months post-deployment**	**t**	**p**	**n**
PTSD symptoms (SRIP) (mean, SD)	26.7 (4.8)	27.4 (6.8)	-2.22	0.03	372
Depressive symptoms (SCL-90 depression subscale) (mean, SD)	17.9 (3.0)	18.2 (4.0)	-1.80	0.07	425
	Pre-deployment	**One year post-deployment**	**t**	**p**	**n**
PTSD symptoms (SRIP) (mean, SD)	26.8 (5.1)	26.8 (6.8)	-0.06	0.95	296
Depressive symptoms (SCL-90 depression subscale) (mean, SD)	18.2 (3.4)	19.0 (5.6)	-2.67	0.008	342
	Pre-deployment	**Two years post-deployment**	**t**	**p**	**n**
PTSD symptoms (SRIP) (mean, SD)	26.7 (5.1)	26.3 (5.6)	1.24	0.21	266
Depressive symptoms (SCL-90 depression subscale) (mean, SD)	17.8 (2.8)	19.0 (5.4)	-3.91	0.0001	328
	Pre-deployment	**Five years post-deployment**	**t**	**p**	**n**
PTSD symptoms (SRIP) (mean, SD)	26.9 (5.2)	28.1 (7.2)	-3.07	0.002	290

4.3.3 Trauma during deployment and childhood trauma

Childhood trauma was significantly associated with PTSD PRSs at $P_T < 0.2$ (beta = 0.29, p = 0.007) after correction for multiple testing, while nominal significant associations were observed at $P_T < 0.1$ (beta = 0.24; p = 0.02) and $P_T < 0.3$ (beta = 0.24, p = 0.02), and at $P_T < 0.1$ for MDD PRSs (beta = 0.26; p=0.02). Trauma during deployment was only associated at nominal significance with MDD PRSs at $P_T < 5 \times 10^{-5}$ (beta = -0.30, p = 0.02).

4.3.4 PTSD symptoms

The development of post-deployment PTSD symptoms at any time point (one month, six months and one, two and five years) was not associated with PTSD PRSs at any P_T. In addition, there were no significant interaction effects of PTSD PRSs with trauma during deployment on the development of PTSD symptoms after deployment (see Table 2, all p-values > 0.02). The added explained variance of PTSD PRSs in the main models was very low or absent (e.g. P_T 0.05 included as main effect, with PTSD symptoms six months post-deployment as the dependent variable: $R^2 = 0.2779$; P_T 0.05 excluded from the model: $R^2 = 0.2772$).

In secondary analyses, PTSD PRSs were not associated with pre-deployment PTSD symptoms (neither direct nor in interaction with childhood trauma; see Supplementary Table 5, all p-values > 0.02). In addition, there were no significant interactions of PTSD PRSs with childhood trauma on the development of post-deployment PTSD symptoms at any post-deployment time point (see Supplementary Table 6, all p-values > 0.01).

4.3.5 Depressive symptoms

No significant main or interaction effects with deployment-related trauma were observed for the MDD PRSs at any P_T on the development of depressive symptoms at any time point after deployment (see Table 3, all p-values > 0.03). Similar to PTSD, MDD PRSs added minimally or not at all to the explained variance of the main models (e.g. P_T 0.05 included as main effect, with depressive symptoms six months post-deployment as the dependent variable: $R^2 = 0.2771843$; P_T 0.05 excluded from the model: $R^2 = 0.2771822$).

In secondary analyses, PRSs for MDD were not associated with baseline depressive symptoms (neither main nor in interaction with childhood trauma; see Supplementary Table 5, all p-values > 0.04). Furthermore, there were no interaction effects of MDD PRSs and childhood trauma on the development of post-deployment depressive symptoms at any time point (see Supplementary Table 7, all p-values > 0.01).

Table 2. PTSD PRSs at all significance thresholds (P_T) in relation to the development of PTSD symptoms after deployment.

P_T PRSs	Total SNPs PTSD	One month post-deployment (N = 421)				Six months post-deployment (N = 372)				One year post-deployment (N = 296)				Two years post-deployment (N = 266)				Five years post-deployment (N = 290)			
		Main effect		PRS x deployment trauma		Main effect		PRS x deployment trauma		Main effect		PRS x deployment trauma		Main effect		PRS x deployment trauma		Main effect		PRS x deployment trauma	
		p	β	p	β	p	β	p	β	p	β	p	β	p	β	p	β	p	β	p	β
$< 5 \times 10^{-6}$	6	0.52	0.1471	0.10	0.1366	0.72	0.1081	0.97	0.0046	0.44	-0.2600	0.42	-0.1071	0.48	-0.2081	0.81	-0.0285	0.23	-0.4469	0.45	0.1096
$< 5 \times 10^{-5}$	59	1.00	-0.0008	0.57	0.0463	0.45	-0.2383	0.70	-0.0405	0.07	-0.6840	**0.02**	**-0.2911**	0.43	-0.2553	0.86	-0.0204	0.43	0.3149	0.38	0.1205
$< 5 \times 10^{-4}$	476	0.89	-0.0326	0.67	0.0326	0.84	0.0627	0.71	-0.0382	0.96	-0.0189	0.54	0.0715	0.99	0.0050	0.42	0.0812	0.16	0.5264	0.07	0.2100
$< 5 \times 10^{-3}$	3,523	0.88	0.0352	0.48	0.0577	0.96	-0.0157	0.83	-0.0227	0.93	-0.0293	0.26	-0.1441	0.88	0.0464	0.50	0.0831	0.62	0.1907	0.98	0.0036
< 0.05	25,514	0.57	0.1387	0.06	0.1485	0.56	-0.1816	0.75	0.0330	0.90	-0.0454	0.47	-0.0843	0.97	-0.0119	0.14	0.1635	0.10	0.6457	0.62	0.0621
< 0.1	44,633	0.84	0.0489	0.20	0.1012	0.79	-0.0837	0.69	0.0426	0.88	0.0564	0.57	-0.0677	0.96	0.0151	0.07	0.2052	0.11	0.6225	0.51	0.0803
< 0.2	76,374	0.82	0.0579	0.19	0.1083	0.51	-0.2119	0.62	0.0530	0.98	-0.0072	0.53	-0.0794	0.81	0.0798	0.08	0.2018	0.09	0.6636	0.61	0.0656
< 0.3	102,562	1.00	-0.0009	0.07	0.1515	0.43	-0.2544	0.33	0.1058	0.83	-0.0765	0.64	-0.0589	0.93	0.0298	0.06	0.2216	0.19	0.5152	0.50	0.0901
< 0.4	125,214	0.98	0.0066	**0.04**	**0.1686**	0.62	-0.1550	0.26	0.1221	0.98	-0.0113	0.95	-0.0088	0.92	0.0342	0.11	0.1877	0.18	0.5307	0.44	0.1028
< 0.5	143,907	0.89	0.0330	**0.04**	**0.1723**	0.63	-0.1525	0.27	0.1188	1.00	0.0005	0.87	-0.0213	0.93	-0.0290	0.11	0.1857	0.20	0.5079	0.52	0.0855
≤ 1	197,108	0.84	0.0499	0.05	0.1627	0.68	-0.1304	0.28	0.1176	0.93	-0.0336	0.89	-0.0175	0.94	-0.0227	0.08	0.2065	0.19	0.5183	0.55	0.0806

Table 3. MDD PRSs at all significance thresholds (P_T) in relation to the development of depressive symptoms after deployment.

P_T PRSs	Total SNPs MDD	One month post-deployment (N = 486)				Six months post-deployment (N = 425)				One year post-deployment (N = 342)				Two years post-deployment (N = 328)			
		Main effect		PRS x deployment trauma		Main effect		PRS x deployment trauma		Main effect		PRS x deployment trauma		Main effect		PRS x deployment trauma	
		p	β	p	β	p	β	p	β	p	β	p	β	p	β	p	β
$< 5 \times 10^{-8}$	45	0.60	-0.0682	0.84	-0.0094	0.07	-0.2896	0.73	-0.0215	0.03	-0.6311	0.94	0.0075	0.51	-0.1805	0.74	0.0348
$< 5 \times 10^{-7}$	101	0.25	-0.1498	0.29	-0.0534	0.09	-0.2684	0.41	-0.0517	0.45	-0.2131	0.70	-0.0410	0.57	0.1566	0.74	0.0350
$< 5 \times 10^{-6}$	198	0.10	-0.2155	0.31	-0.0461	0.11	-0.2573	0.87	-0.0095	0.82	0.0654	0.73	-0.0348	0.90	-0.0345	0.94	-0.0077
$< 5 \times 10^{-5}$	552	0.95	0.0081	0.92	-0.0045	0.93	-0.0133	0.71	-0.0217	0.11	0.4659	0.22	-0.1323	0.22	0.3387	0.38	-0.0930
$< 5 \times 10^{-4}$	1,821	0.38	-0.1132	0.83	-0.0097	0.42	-0.1258	0.49	-0.0406	0.99	0.0020	0.24	-0.1244	0.59	0.1498	0.42	-0.0844
$< 5 \times 10^{-3}$	7,221	0.40	-0.1084	0.34	-0.0419	0.53	-0.1007	0.52	-0.0355	0.30	-0.2863	0.09	-0.1712	0.78	0.0746	0.98	0.0022
< 0.05	31,728	0.33	-0.1289	0.72	-0.0165	0.85	-0.0304	0.94	-0.0041	0.69	0.1226	0.38	-0.1014	0.61	0.1448	0.82	-0.0243
< 0.1	49,995	0.16	-0.1891	0.66	-0.0203	0.73	-0.0581	0.87	0.0100	0.94	0.0213	0.33	-0.1133	0.60	0.1514	0.83	-0.0231
< 0.2	78,345	0.11	-0.2112	0.55	-0.0286	0.59	-0.0889	0.84	0.0121	0.97	0.0115	0.13	-0.1755	0.44	0.2192	0.79	-0.0291
< 0.3	100,527	0.11	-0.2110	0.68	-0.0196	0.67	-0.0696	0.70	0.0231	0.77	0.0893	0.18	-0.1551	0.43	0.2266	0.88	-0.0167
< 0.4	118,804	0.19	-0.1759	0.97	-0.0020	0.79	-0.0431	0.52	0.0389	0.88	0.0445	0.17	-0.1589	0.51	0.1895	0.94	0.0084
< 0.5	133,783	0.19	-0.1764	0.92	-0.0049	0.94	-0.0121	0.56	0.0352	0.83	0.0672	0.22	-0.1432	0.62	0.1435	0.95	0.0074
≤ 1	176,024	0.20	-0.1709	0.94	-0.0035	0.92	-0.0166	0.59	0.0327	0.92	0.0326	0.23	-0.1395	0.74	0.0938	0.99	-0.0020

4.4 Discussion

In a large prospective military cohort, we calculated PRSs derived from recent GWASs on PTSD and MDD to investigate both main effects of genetic vulnerability and the interaction with deployment-related trauma on the development of self-reported PTSD- and depression-related symptoms at five time points over five years after deployment to Afghanistan. Meta-analysis of GWAS results for two PTSD studies and two MDD studies allowed us to achieve substantial power for PRS calculation. In secondary analyses, we examined the association of these PRSs with baseline depressive and PTSD symptoms and their interaction with childhood trauma. We did not find significant main or interaction effects of PRSs on PTSD or depressive symptoms at any time point after deployment. In addition, there were no significant associations with baseline symptom levels and no interactions with childhood trauma on pre- and the development of post-deployment symptoms levels.

The absence of evidence of a relation between PTSD PRSs and our outcomes may have several explanations. First, the SNP-based heritability in the European descent subsamples of the original studies was either not significant [8] or only significant in females [6] while our target cohort largely consisted of males, thus indicating that our PRSs may have been insufficiently refined to discover significant associations. Second, the discovery GWASs [6,8] contained mostly trauma-exposed controls, thereby excluding genetic variation associated with liability to exposure to trauma. Exposure to trauma is itself moderately heritable [23], which is possibly mediated by specific personality traits (e.g. harm avoidance and/or novelty seeking). As such, the PTSD PRSs do not capture much of the genetic variation associated with PTSD through specific personality traits, whereas our military cohort was not selected on these traits. This may have contributed to a difference in genetic background of PTSD in the discovery samples [6,8] compared with our target sample. Nonetheless, the association of PTSD PRSs with childhood trauma at one P_T suggests that at least some of the genetic risk captured by the PRSs could reflect personality traits leading to more trauma exposure. Lastly, PTSD in part of the discovery sample had been caused by pre-deployment or non-military traumatic events [6,8]. In the STARRS cohort, only 24% of the soldiers had been deployed and most participants were in the age range of 18-20 years [8]. Of note, PTSD caused by childhood trauma

may have a different genetic etiology than PTSD caused by deployment-related trauma.

The lack of significant findings with MDD PRSs is noteworthy given the results of two previous studies confirming the validity of an MDD PRS based on a maximum of 9,240 cases and 9,519 controls [24] in two different populations [25,26]. Peyrot *et al.* [26] found both main and interaction effects of the MDD PRS and childhood trauma on risk for MDD in a study sample of 1,645 patients with MDD and 340 controls. In addition, Musliner *et al.* [25] found a main effect of the MDD PRS and stressful life events on depressive symptoms in a population of 8,761 mainly older adults. The MDD PRSs in the present study were based on over 135,000 cases and 344,000 controls. This would expectedly lead to more valid PRSs than the ones based on the previous, much smaller GWAS [24], even though depressive symptoms in part of the discovery data [9] were ascertained using self-report. In contrast to PTSD, where male SNP based heritability was low and non-significant, the SNP-based heritability in the MDD GWAS [7] was significant and estimated at 8.7%. Possibly, the null finding in the present study, with very low or no added explained variances of the PRSs in our main models, was due to etiological differences in depressive phenotypes of the military cohort after deployment versus patients with a depressive episode in the general population, the minor increase in depressive symptoms between the pre-deployment and post-deployment measurements and the relatively low level of severe depression in our study sample. More specific phenotyping of heterogeneous disorders such as MDD and PTSD in both the discovery and target cohorts may be effective in this regard, as some studies have shown that this increases power to find genome-wide significant associations [27,28]. Furthermore, genetic correlation estimates could provide insight into genetic heterogeneity between discovery and target datasets when sample sizes are sufficient and future methodological studies could focus on loss of power in PRS analysis due to phenotypic and etiological heterogeneity.

The longitudinal design around military deployment, with psychopathology outcomes at six time points and quantification of deployment-related trauma, is the main strength of this study. The relatively healthy population represents this study's most important limitation. The increase in PTSD and depressive symptoms after deployment was small and at some time points non-significant,

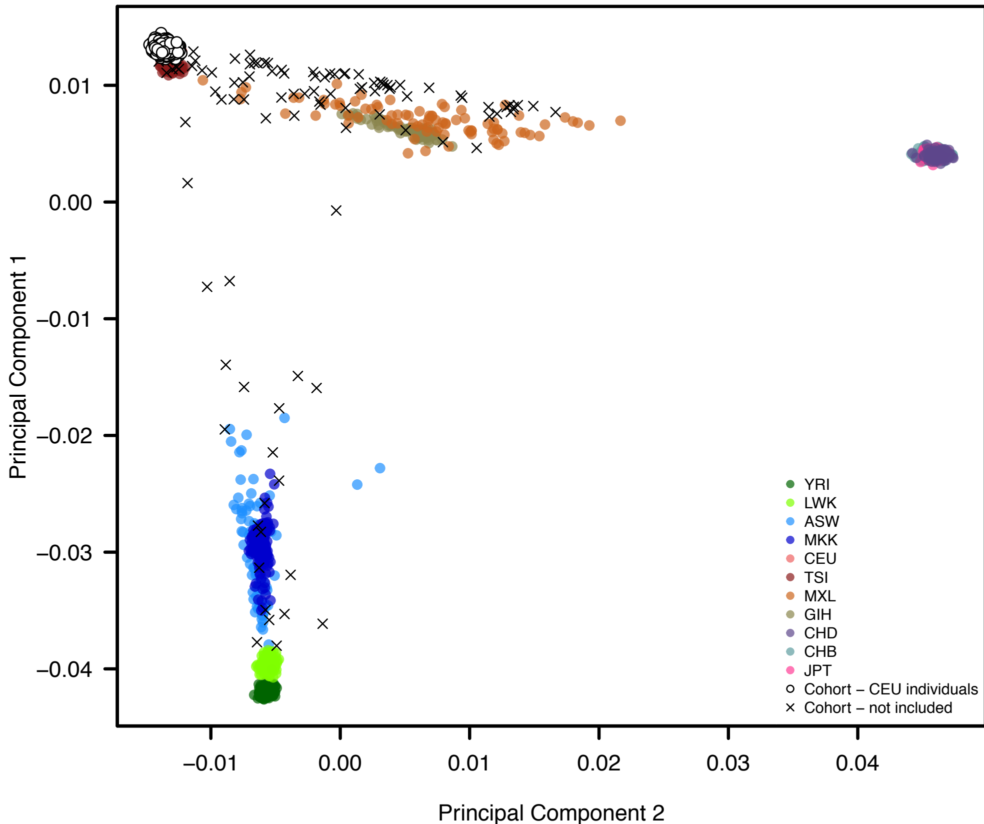
probably hampering the power in our analyses even with large discovery GWASs on which PRSs were calculated. Moreover, mean pre- and post-deployment scores do not fall in the range of severe psychopathology for PTSD (SRIP ≥ 38) [29] and above-average depression. Furthermore, self-report questionnaires were used to evaluate symptom severity, which may have led to some social desirability bias. Moreover, trauma exposure questionnaires did not take subjective experience of these events into account, while this subjective experience is probably more important than the actual event [30]. In addition, the present sample size may have been insufficient. This issue, however, is partly compensated for by the longitudinal design. Lastly, as mentioned above, the PRSs of PTSD were based on a relatively small GWAS (as compared to the MDD discovery dataset), compromising their validity.

Future studies may further shed light upon the validity of MDD PRSs in non-clinical and clinical samples. For PTSD, a larger upcoming GWAS by the Psychiatric Genomics Consortium (PGC) [31] may yield more refined PTSD PRSs than the ones used here. In addition to increasing sample sizes, the focus in GWASs may need to be shifted towards more specific [27] and/or more severe [28] phenotypes. As such, PRSs may be most informative about etiology and may ultimately be useful for prediction purposes [32].

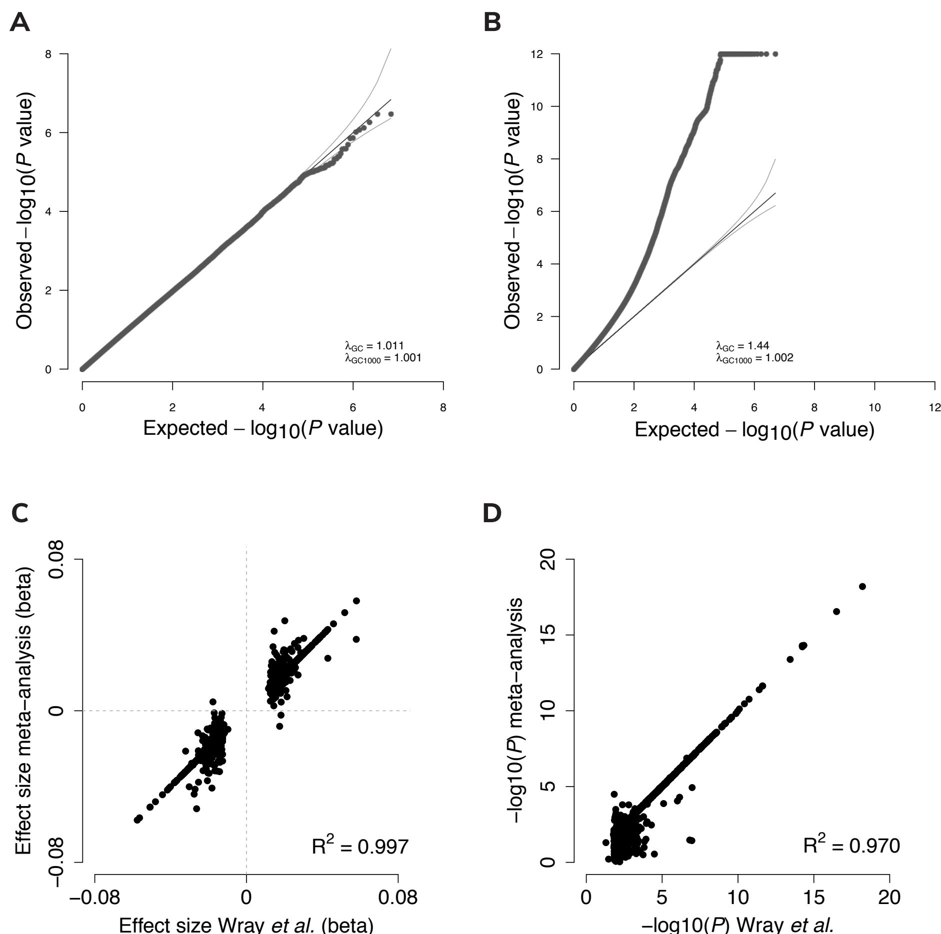
In conclusion, we used polygenic risk scores derived from large GWASs for PTSD and MDD to investigate the interaction of genetic vulnerability with potentially traumatic events on the development of these disorders in a prospective military cohort deployed to a combat zone. Our results indicate the limited validity of PTSD and MDD PRSs in this relatively healthy military population and highlight the importance of etiological heterogeneity of these disorders. Future studies investigating common genetic risk for PTSD will benefit from increased sample sizes in GWASs, but for both MDD and PTSD phenotype refinement may be of crucial importance for progress in this field.

4.5 Supplementary information

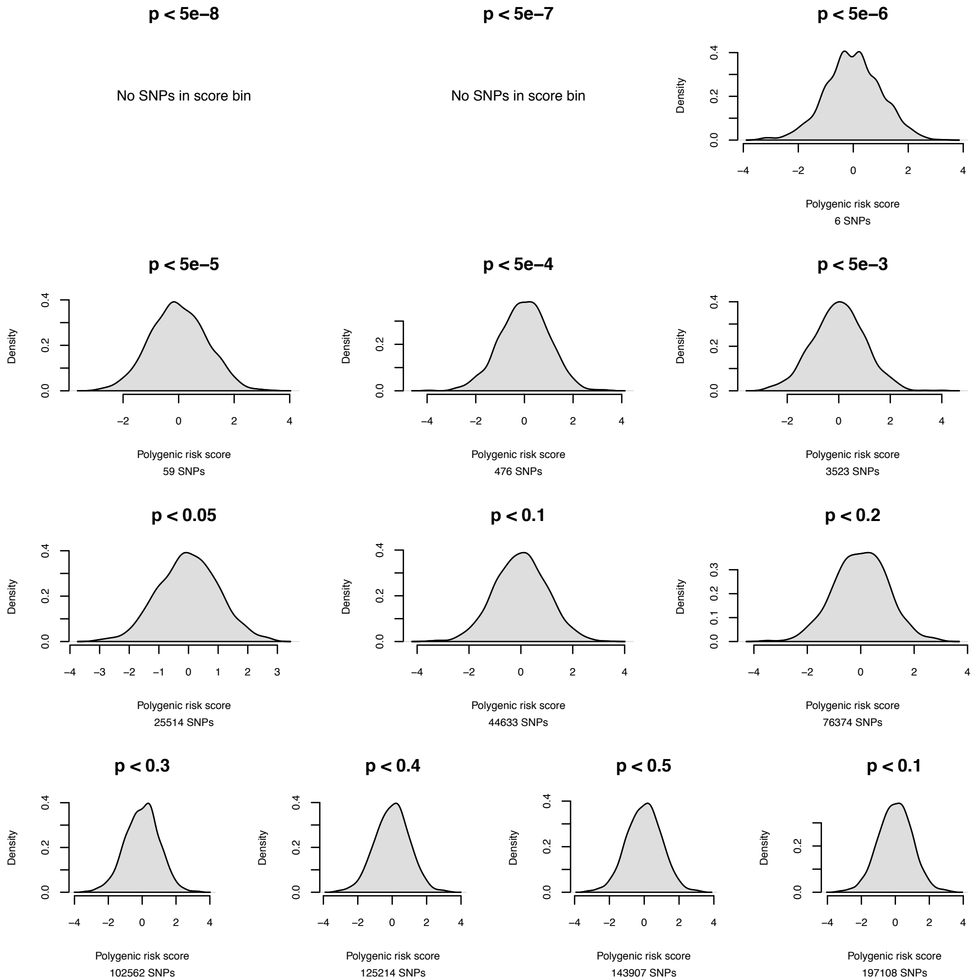
4.5.1 Supplementary figures



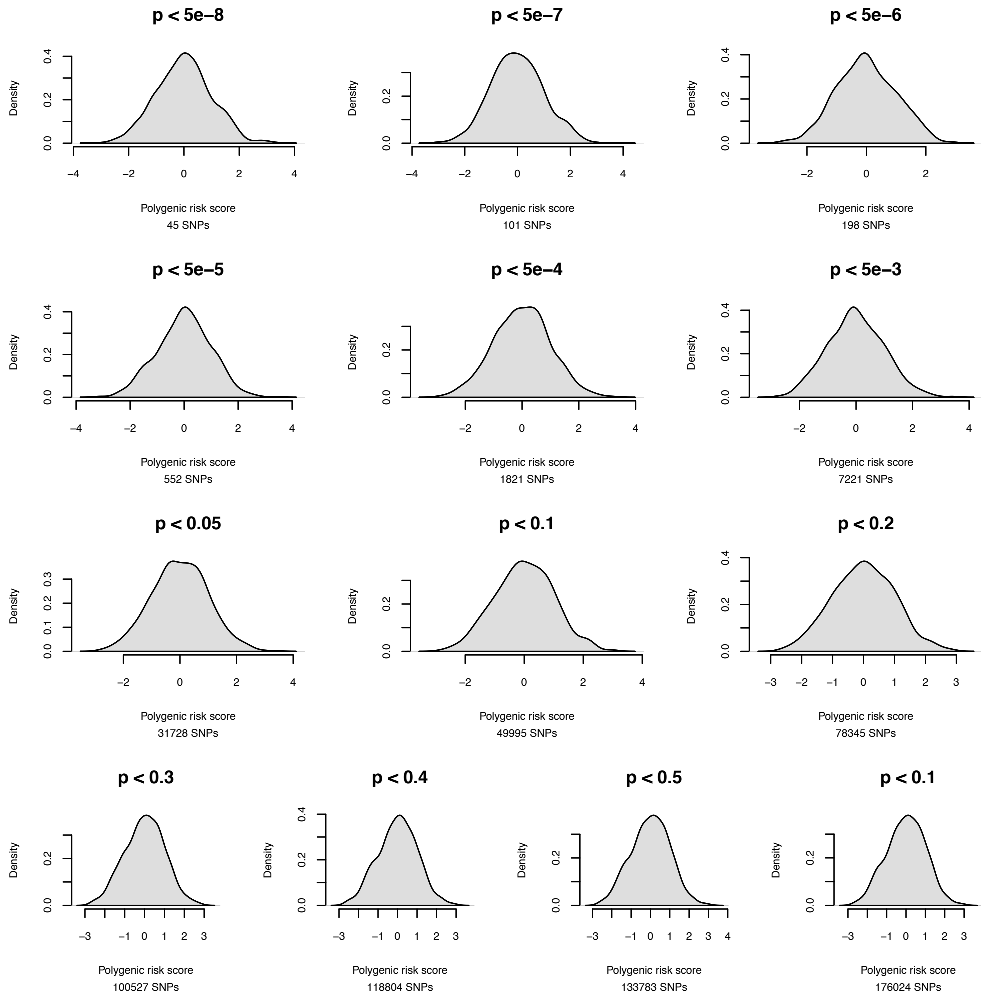
Supplementary Figure 1. Population structure of the cohort and selection of individuals with a Central-European ancestry. Principal components were calculated from the genetic data and the first two principal components were plotted on a HapMap3 background. Colored dots indicate HapMap3 individuals, white dots indicate individuals with a Central-European ancestry from our cohort that were selected for polygenic risk score analyses, and crosses indicate individuals in our cohort that did not have a Central-European genetic ancestry and were excluded from analyses. HapMap3 populations: CEU, Utah residents with Northern and Western European ancestry; Yoruba in Ibadan, Nigeria (YRI); Luhya in Weyube, Kenya (LWK); African ancestry in Southwest USA (ASW); Maasai in Kinyawa, Kenya (MKK); Utah residents with Northern and Western European ancestry (CEU); Toscani in Italy (TSI); Mexican ancestry in Los Angeles, California (MXL); Gujarati Indians in Houston, Texas (GIH); Chinese in Metropolitan Denver, Colorado (CHD); Han Chinese in Beijing, China (CHB); Japanese in Tokyo, Japan (JPT).



Supplementary Figure 2. Meta-analysis results for PTSD and depression GWAS datasets. Quantile-quantile (QQ) plots show observed versus expected $-\log_{10}$(p-value) distributions for the PTSD meta-analysis using data from of Stein *et al.* (2016) and Duncan *et al.* (2017) (A), and depression using the available results for depression from Wray *et al.* (2018) and the full result from Hyde *et al.* (2015) (association p-values < 10^{-12} were truncated at p = 10^{-12}) (B). λ_{GC} indicates the inflation of p-values from the expected null, and λ_{GC1000} the inflation scaled to a study of 1000 cases and 1000 controls. The QQ plot in (B) replicates the full result from Wray *et al.* (2018) with similar λ_{GC1000} (1.002) [7]. Robustness of the recovered full result of Wray *et al.* 2018 is supported by high correlations between effect sizes (0.997) (C) and $-\log_{10}$(p-values) (0.97) (D) for 8,398 LD-independent top SNPs for which full data was shared (www.med.unc.edu/pgc/results-and-downloads, under MDD2 2018).



Supplementary Figure 3. Distribution of PTSD polygenic risk scores. Density plots show the distribution of PTSD PRSs per p-value threshold (P_T). The number of SNPs used for PRS calculation in each P_T is shown under each plot. No scores were calculated for the two most stringent P_T as no SNPs were available. PRSs for all P_Ts were scaled around mean zero with a variance of one.



Supplementary Figure 4. Distribution of MDD polygenic risk scores. Density plots show the distribution of MDD PRSs per p-value threshold (P_T). The number of SNPs used for PRS calculation in each P_T is shown underneath each plot. PRSs for all P_Ts were scaled around mean zero with a variance of one.

4.5.2 Supplementary tables

For **Supplementary Tables 1 and 2**, please refer to Supplementary Tables 2 and 3 in Chapter 3.

Supplementary Table 3. Genomic regions with high LD removed prior to polygenic risk score calculation.

Chromosome	Basepair start	Basepair end
1	48000000	52000000
2	86000000	100500000
	183000000	190000000
3	47500000	50000000
	83500000	87000000
5	44500000	50500000
	129000000	132000000
6	25500000	33500000
	57000000	64000000
	140000000	142500000
7	55000000	66000000
8	8000000	12000000
	43000000	50000000
	8135000	12000000
	112000000	115000000
10	37000000	43000000
11	87500000	90500000
12	33000000	40000000
17	40900000	45000000
20	32000000	34500000

Positions refer to human genome build GRCh37

Supplementary Table 4. Bivariate LD Score regression intercepts for meta-analyzed datasets.

GWAS 1	GWAS 2	r_g	Bivariate intercept (SE)
Depression (Hyde *et al.* 2016)	Depression (Wray *et al.* 2018)*	0.86 (0.03)	0.0038 (0.0058)
PTSD (Stein *et al.* 2016) NSS1 cohort	PTSD (Duncan *et al.* 2017)	NA	0.0047 (0.004)
PTSD (Stein *et al.* 2016) NSS2 cohort	PTSD (Duncan *et al.* 2017)	NA	-0.0009 (0.0844)
PTSD (Stein *et al.* 2016) PPDS cohort	PTSD (Duncan *et al.* 2017)	NA	0.0004 (0.0039)
PTSD (Stein *et al.* 2016) NSS1 cohort	PTSD (Stein *et al.* 2016) NSS2 cohort	NA	-0.0103 (0.005)
PTSD (Stein *et al.* 2016) NSS1 cohort	PTSD (Stein *et al.* 2016) PPDS cohort	NA	0.0016 (0.0041)
PTSD (Stein *et al.* 2016) NSS2 cohort	PTSD (Stein *et al.* 2016) PPDS cohort	NA	0.0040 (0.0041)

r_g is genetic correlation, calculated only between the two depression datasets as low sample size of PTSD GWAS cohorts (< 5,000) did not allow us to calculate reliable estimates of genetic correlation.

* GWAS summary statistics as shared on the website of the Psychiatric Genomics Consortium (www.med.unc.edu/pgc/results-and-downloads, under MDD2 2018). These data do not include the Hyde *et al.* results.

Supplementary Table 5. PRSs at all significance thresholds (P_T) in relation to pre-deployment PTSD and depressive symptoms.

P_T PRSs	Total SNPs PTSD	PTSD PRS in relation to baseline PTSD symptoms (N = 590)				Total SNPs MDD	MDD PRS in relation to baseline depressive symptoms (N = 705)			
		Main effect		PRS x early trauma			Main effect		PRS x early trauma	
		p	β	p	β		p	β	p	β
< 5 x 10^{-8}	0	NA	NA	NA	NA	45	0.32	0.1154	0.25	-0.0456
< 5 x 10^{-7}	0	NA	NA	NA	NA	101	**0.04**	**0.2437**	0.06	-0.0702
< 5 x 10^{-6}	6	0.47	-0.1536	0.34	-0.0802	198	0.19	0.1545	0.36	-0.0365
< 5 x 10^{-5}	59	0.71	0.0822	0.59	0.0458	552	0.69	0.0467	0.21	-0.0531
< 5 x 10^{-4}	476	0.32	0.2143	0.77	-0.0259	1,821	0.27	-0.1272	0.60	-0.0227
< 5 x 10^{-3}	3,523	0.92	-0.0219	0.88	0.0126	7,221	0.60	0.0601	0.38	-0.0360
< 0.05	25,514	0.18	0.2906	0.18	0.1100	31,728	0.38	0.1018	0.14	-0.0599
< 0.1	44,633	0.05	0.4230	0.09	0.1514	49,995	0.08	0.2014	0.42	-0.0343
< 0.2	76,374	**0.02**	**0.4864**	**0.03**	**0.2003**	78,345	0.34	0.1109	0.31	-0.0426
< 0.3	102,562	**0.02**	**0.4882**	0.14	0.1281	100,527	0.42	0.0945	0.25	-0.0491
< 0.4	125,214	**0.04**	**0.4327**	0.26	0.0963	118,804	0.44	0.0910	0.23	-0.0516
< 0.5	143,907	**0.04**	**0.4352**	0.21	0.1067	133,783	0.45	0.0888	0.27	-0.0474
≤ 1	197,108	**0.03**	**0.4525**	0.13	0.1307	176,024	0.46	0.0870	0.28	-0.0462

Supplementary Table 6. The interaction of PTSD PRSs at all significance thresholds (P_T) with childhood trauma on the development of PTSD symptoms after deployment.

		One month post-deployment (N = 421)		Six months post-deployment (N = 372)		One year post-deployment (N = 296)		Two years post-deployment (N = 266)		Five years post-deployment (N = 290)	
		PRS x childhood trauma		PRS x childhood trauma		PRS x childhood trauma		PRS x childhood trauma		PRS x childhood trauma	
P_T PRSs	Total SNPs PTSD	p	β	p	β	p	β	p	β	p	β
$< 5 \times 10^{-6}$	6	0.66	0.0438	0.18	0.1714	0.78	-0.0409	0.72	-0.0425	0.66	0.0679
$< 5 \times 10^{-5}$	59	0.55	-0.0580	0.18	0.1628	0.93	0.0139	0.85	-0.0253	0.06	0.2979
$< 5 \times 10^{-4}$	476	0.80	-0.0257	0.42	0.0987	0.75	0.0455	0.70	0.0458	***0.01***	**0.3862**
$< 5 \times 10^{-3}$	3,523	0.18	0.1365	**0.03**	**0.2796**	0.10	0.2526	**0.03**	**0.2829**	***0.01***	**0.4558**
< 0.05	25,514	0.94	-0.0081	0.40	0.1079	0.40	0.1181	0.10	0.1997	***0.02***	**0.3718**
< 0.1	44,633	0.70	0.0434	0.33	0.1353	0.53	0.0962	0.22	0.1601	***0.04***	**0.3479**
< 0.2	76,374	0.36	0.1071	0.39	0.1257	0.40	0.1426	0.26	0.1601	0.10	0.2992
< 0.3	102,562	0.58	0.0587	0.39	0.1152	0.28	0.1709	0.18	0.1772	0.11	0.2754
< 0.4	125,214	0.77	0.0308	0.40	0.1095	0.46	0.1122	0.17	0.1735	0.11	0.2629
< 0.5	143,907	0.63	0.0506	0.27	0.1411	0.41	0.1247	0.08	0.2209	0.06	0.3086
≤ 1	197,108	0.65	0.0481	0.27	0.1425	0.41	0.1272	0.09	0.2175	0.08	0.2884

4

Supplementary Table 7. The interaction of MDD PRSs at all significance thresholds (P_T) with childhood trauma on the development of depressive symptoms after deployment.

		One month post-deployment (N = 486)		Six months post-deployment (N = 425)		One year post-deployment (N = 342)		Two years post-deployment (N = 328)	
		PRS x childhood trauma		PRS x childhood trauma		PRS x childhood trauma		PRS x childhood trauma	
P_T PRSs	Total SNPs MDD	p	β	p	β	p	β	p	β
$< 5 \times 10^{-8}$	45	0.44	-0.0371	0.60	-0.0303	0.41	-0.0860	0.10	-0.1483
$< 5 \times 10^{-7}$	101	0.74	-0.0154	0.70	0.0216	0.22	0.1219	0.94	-0.0068
$< 5 \times 10^{-6}$	198	0.47	-0.0325	0.84	-0.0103	0.99	-0.0015	0.59	-0.0450
$< 5 \times 10^{-5}$	552	0.65	0.0236	0.50	0.0410	0.20	0.1427	0.08	0.1831
$< 5 \times 10^{-4}$	1,821	0.67	0.0236	0.89	0.0089	0.13	0.1710	0.47	0.0733
$< 5 \times 10^{-3}$	7,221	0.86	0.0088	0.27	0.0651	**0.01**	**0.2597**	0.67	0.0425
< 0.05	31,728	0.54	0.0307	0.37	0.0527	**0.05**	**0.2330**	0.15	0.1491
< 0.1	49,995	0.79	0.0139	0.70	0.0229	0.23	0.1460	0.22	0.1296
< 0.2	78,345	0.68	0.0207	0.73	0.0202	0.31	0.1154	0.15	0.1428
< 0.3	100,527	0.72	0.0186	0.60	0.0306	0.21	0.1400	0.18	0.1367
< 0.4	118,804	0.75	0.0167	0.75	0.0196	0.16	0.1651	0.21	0.1321
< 0.5	133,783	0.72	0.0189	0.77	0.0178	0.16	0.1614	0.21	0.1325
≤ 1	176,024	0.68	0.0212	0.78	0.0168	0.14	0.1698	0.18	0.1423

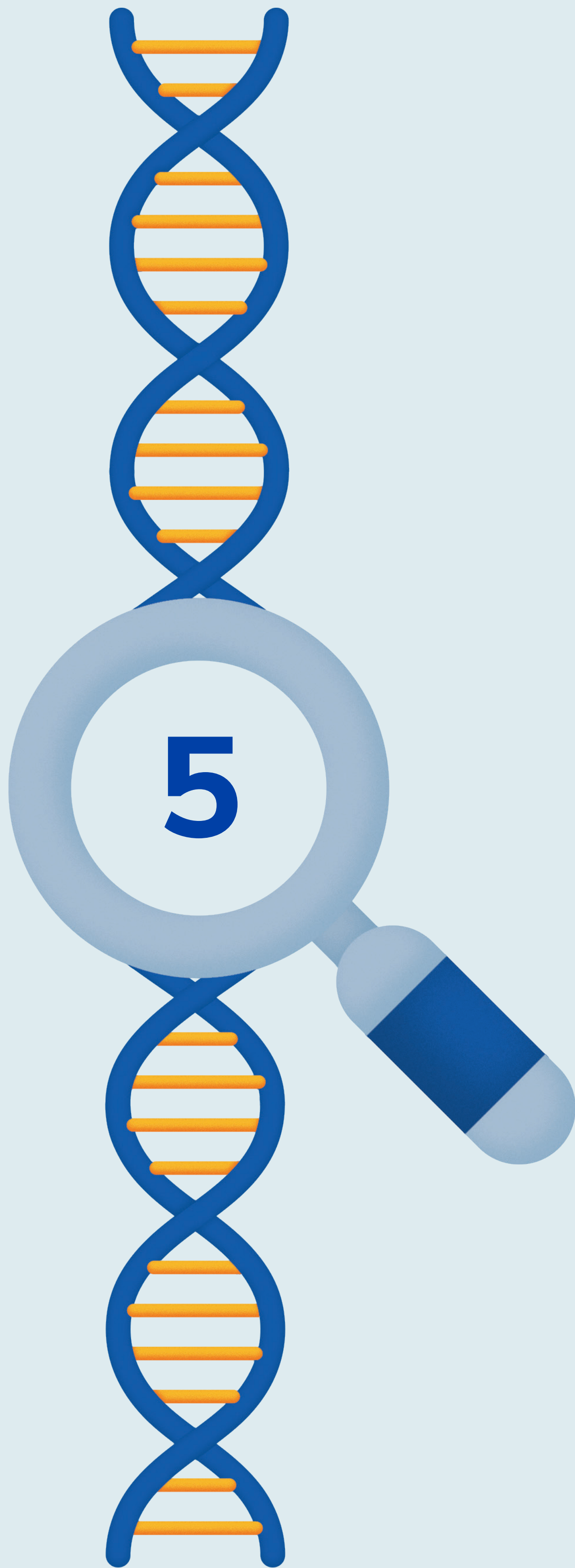
4.6 References

1. Hoge CW, Auchterlonie JL, Milliken CS. Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. JAMA 2006;295(9):1023–32.
2. Reijnen A, Rademaker AR, Vermetten E, Geuze E. Prevalence of mental health symptoms in Dutch military personnel returning from deployment to Afghanistan: a 2-year longitudinal analysis. Eur Psychiatry 2015;30(2):341–6.
3. Karam EG, Friedman MJ, Hill ED, Kessler RC, McLaughlin KA, Petukhova M, et al. Cumulative traumas and risk thresholds: 12-month PTSD in the World Mental Health (WMH) surveys. Depress Anxiety 2014;31(2):130–42.
4. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. Arch Neurol 2005;62(6):617.
5. Kok BC, Herrell RK, Thomas JL, Hoge CW. Posttraumatic stress disorder associated with combat service in Iraq or Afghanistan: reconciling prevalence differences between studies. J Nerv Ment Dis 2012;200(5):444–50.
6. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. Mol Psychiatry 2018;23(3):666–73.
7. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018;50(5):668–81.
8. Stein MB, Chen CY, Ursano RJ, Cai T, Gelernter J, Heeringa SG, et al. Genome-wide Association Studies of Posttraumatic Stress Disorder in 2 Cohorts of US Army Soldiers. JAMA Psychiatry 2016;
9. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet 2016;48(9):1031–6.
10. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research Review: Polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry 2014;55(10):1068–87.
11. van Zuiden M, Geuze E, Willemen HLDM, Vermetten E, Maas M, Heijnen CJ, et al. Pre-existing high glucocorticoid receptor number predicting development of posttraumatic stress symptoms after military deployment. Am J Psychiatry 2011;168(1):89–96.
12. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48(10):1284–7.
13. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48(10):1279–83.
14. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 2016;48(11):1443–8.
15. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56–65.
16. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.
17. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190–1.
18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4(1):7.
19. Hovens JE, Bramsen I, van der Ploeg HM. Self-rating inventory for posttraumatic stress disorder: review of the psychometric properties of a new brief Dutch screening instrument. Percept Mot Skills 2002;94(3 Pt 1):996–1008.

20. Arrindell WA, Ettema J. Revised manual for a multidimensional indicator of psychopathology In: Herziene handleiding bij een multidimensionele psychopathologie indicator. Swets Zeitlinger, Lisse, the Netherlands 2003;
21. Bremner JD, Bolus R, Mayer EA. Psychometric properties of the Early Trauma Inventory-Self Report. J Nerv Ment Dis 2007;195(3):211–8.
22. Keller MC. Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. Biol Psychiatry 2014;75(1):18–24.
23. True WR, Rice J, Eisen SA, Heath AC, Goldberg J, Lyons MJ, et al. A twin study of genetic and environmental contributions to liability for posttraumatic stress symptoms. Arch Neurol 1993;50(4):257–64.
24. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013;18(4):497–511.
25. Musliner KL, Seifuddin F, Judy JA, Pirooznia M, Goes FS, Zandi PP. Polygenic risk, stressful life events and depressive symptoms in older adults: a polygenic score analysis. Psychol Med 2015;45(8):1709–20.
26. Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, et al. Effect of polygenic risk scores on depression in childhood trauma. Br J Psychiatry 2014;205(2):113–9.
27. Milaneschi Y, Lamers F, Peyrot WJ, Abdellaoui A, Willemsen G, Hottenga JJ, et al. Polygenic dissection of major depression clinical heterogeneity. Mol Psychiatry 2016;21(4):516–22.
28. CONVERGEconsortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature 2015;523(7562):588–91.
29. van Zelst WH, de Beurs E, Beekman ATF, Deeg DJH, Bramsen I, van Dyck R. Criterion validity of the self-rating inventory for posttraumatic stress disorder (SRIP) in the community of older adults. J Affect Disord 2003;76(1-3):229–35.
30. Conway CC, Hammen C, Espejo EP, Wray NR, Najman JM, Brennan PA. Appraisals of Stressful Life Events as a Genetically-Linked Mechanism in the Stress–Depression Relationship. Cognit Ther Res 2011;36(4):338–47.
31. Nievergelt CM, Ashley-Koch AE, Dalvie S, Hauser MA, Morey RA, Smith AK, et al. Genomic Approaches to Posttraumatic Stress Disorder: The Psychiatric Genomic Consortium Initiative. Biol Psychiatry 2018;83(10):831–9.
32. Smoller JW. The Genetics of Stress-Related Disorders: PTSD, Depression, and Anxiety Disorders. Neuropsychopharmacology 2016;41(1):297–319.

Part 2

Cross-disorder studies


5

Chapter 5

Genetic cross-disorder analysis in psychiatry: from methodology to clinical utility

---

Dick Schijven, Jan H. Veldink, Jurjen J. Luykx

In press at *The British Journal of Psychiatry*

Abstract

Genome-wide association studies (GWASs) have uncovered hundreds of loci associated with psychiatric disorders. Cross-disorder studies are among the prime ramifications of such research. Here, we discuss the methodology of the most widespread methods and their clinical utility with regard to diagnosis, prediction, disease etiology and treatment in psychiatry.

5.1 Introduction

In the field of psychiatric genetics, thousands of affected and healthy individuals have been included in genome-wide association studies (GWASs) to test disease associations at common genetic variants, of which single-nucleotide polymorphisms (SNPs) are most abundant. These occur throughout the genome on average once every couple of hundreds of base pairs and have minor allele frequencies > 1% in the general population. Currently, arrays are available to genotype over two million SNPs simultaneously. Genotypes of SNPs that lie close together in the genome are highly likely to be co-inherited and therefore show high correlation, resulting in blocks of SNPs that are in linkage disequilibrium (LD). LD can be exploited to impute SNPs that were not directly genotyped on an array, at present increasing the number of SNPs in GWAS datasets up to 10-15 million variants. Moreover, an association signal usually covers a genomic region of variable size, called a locus, wherein the associated variants can be in LD with one or more causal variant(s). Most SNPs are known to have small effect sizes in mental illness (odds ratios between 1.03-1.10) [1,2], but they collectively account for about 20-30% of the heritability per disorder (SNP-based heritability) [3], of which only a fraction is currently explained by genome-wide significant SNPs discovered in GWASs (association p-value < 5×10^{-8}), implying that a plethora of SNPs which have not yet been discovered by GWASs due to insufficient power underlies these "complex polygenic" traits [1,4]. Interestingly, many loci are not exclusively associated with a single disorder but with two or even more, suggesting etiological similarities between diseases. This opens up opportunities to evaluate the shared genetic basis of psychiatric disorders. Thus, SNP data are the starting point for many cross-disorder analyses and are usually available as either individual-level datasets or post-GWAS summary-level data. Individual-level data include SNP genotypes, phenotypes and other relevant information for each participant separately. Summary-level data (or summary statistics) comprise the final results of a GWAS, including per-SNP information on their association with the trait studied (usually odds ratios and p-values). Summary-level data are easier to obtain through public sharing (e.g. www.ebi.ac.uk/gwas/downloads/summary-statistics) and computationally easier to handle, at the cost of less detailed information that come with individual-level data. Cross-disorder studies aim to capture both pleiotropy and genetic correlation, which are related

but essentially different concepts. Pleiotropy defines a single genetic variant affecting more than one phenotype, whereas genetic correlation describes the sharing of polygenic disease architectures at a genome-wide level. Cross-disorder studies can be utilized in several clinical domains, namely: 1) diagnostics (disorder classification and patient stratification), 2) prognosis (prediction of clinical course and outcome of a disorder), and 3) treatment (discovery of drug targets and treatment tailoring through new insights into disease etiology). Here, we outline the current and potential future contributions of widely applied approaches in genetic cross-disorder studies for the field of psychiatry, ordered by the clinical domain they target (also see Table 1).

5.2 Diagnostics

Genetic correlations are the most widely used estimate to describe the extent to which variance in liability for two disorders is attributable to additive genetic effects and thus expresses the proportion of shared SNP-based heritability between two traits. Initially, methods to estimate genetic correlation from GWAS data relied on individual-level genotype data (e.g. bivariate restricted maximum likelihood, REML), using a genetic relationship matrix that captures genetic similarity between distantly related individuals which can be correlated to phenotypic similarities for two diseases [5]. This was followed by methods using summary-level data, where linkage disequilibrium score regression (LDSC) is widely applied and genetic correlations are based on similarities in effect sizes and effect directions of SNPs that are shared between GWAS summary statistics of two phenotypes [6]. Bivariate REML and especially LDSC are most accurate for traits with a polygenic architecture, rendering both well applicable to psychiatric disorders. The genetic correlation estimate falls within the range of -1 and 1, where 1 represents the unlikely scenario that shared liability is caused by exactly the same risk SNPs and -1 by the similarly implausible situation where exactly identical SNPs increase risk for one disease while decreasing risk for another. A null correlation indicates absence of shared disease liability due to overlapping genetic architecture. Genetic correlation estimates are based on all SNPs present in two GWAS datasets or summary-level results regardless of their strength of association with a disease (as sub-threshold variants capture a substantial amount of the SNP-based heritability). The ability to test these correlations at large scale has provided

insights into the landscape of disease classifications both within psychiatric disorders as well as for psychiatric disorders relative to other disease groups. Significant genetic correlations are often observed between psychiatric disorders, with schizophrenia and bipolar disorder (BD) peaking at approximately 0.8 [3,6,7]. Surprising genetic correlations between psychiatric traits and other phenotypes presumed to be etiologically unrelated have been described, suggesting that biological mechanisms involved in psychiatric disorders overlap with those of phenotypes that are clinically distinct. Examples include significant positive genetic correlations between BMI on the one hand and major depressive disorder (MDD) (0.11) and attention-deficit hyperactivity disorder (ADHD) (0.21) on the other [7]. The positive genetic correlation between schizophrenia and amyotrophic lateral sclerosis (ALS) of 0.14 is the first described genome-wide correlation between a psychiatric and a neurological disorder [8].

Polygenic risk score (PRS) analysis is another widely applied method to assess polygenic overlap between diseases. This technique uses estimates of SNPs derived from summary-level data of a discovery GWAS to calculate per-individual scores based on the number of effect alleles carried and weighted for the effect sizes at all overlapping SNPs in an individual-level target GWAS dataset. It is possible to calculate these scores on all SNPs tested for association in the discovery GWAS ($p \leq 1$), but a selection is often made based on strength of association (e.g. $p < 5 \times 10^{-8}$). These scores thus capture the combined effects of many SNPs and in cross-disorder studies PRSs reflecting the polygenic risk for a discovery disease are tested for association with the phenotypic measure of a different target disease [9]: for example, schizophrenia PRSs explain up to 2.4% of the phenotypic variance in BD [10,11]. Strikingly, schizophrenia PRSs were more strongly associated with schizoaffective BD, followed by BD type I and BD type II (risk ratios of 1.37, 1.30 and 1.04, respectively), showing that etiological differences exist within psychiatric disorders [12]. It is furthermore particularly interesting to compare individuals at the extreme ends of the range of PRSs in a study population. For instance, when PRSs calculated using summary-level schizophrenia GWAS data were applied to an independent schizophrenia case-control target GWAS dataset, individuals in the highest PRS decile had up to 20-fold increased odds of schizophrenia compared to individuals in the lowest PRS decile [13], again showing the substantial combined

effect of many SNPs. In addition to this within-trait example, the observation of increased odds of disease is to a lesser extent applicable to correlated diseases in between-disorder studies (e.g. highest decile schizophrenia PRSs increasing odds of ALS up to 1.3 compared to lowest decile PRSs in a case-control cohort) [8]. Explained variances based on PRSs are always much lower than genetic correlation estimates, which has been demonstrated for schizophrenia and BD (2.4% vs. 80%). These methods are conceptually different: while genetic correlations describe the proportion of shared genetic background based on theoretical full SNP-based heritabilities, PRSs are based on actual SNP effect estimates and applied to explain phenotypic variance that is only partly attributable to genetic factors. PRSs can thus never explain more variance than the heritability of the target disease or, when applied in cross-disorder analysis, the proportion of the heritability shared with the discovery disease [14].

As with genetic correlations, PRSs can be useful to establish genetic links between disorders and define the genetic landscape of psychiatric disorders. In a population of unselected samples and when calculated based on a large number of SNPs, PRSs show a normal distribution. Therefore, the true diagnostic utility of PRSs as a quantitative phenotype lies in disease classification and the more fine-grained stratification of patients. For example, BD with mood-incongruent psychosis shows a stronger correlation to schizophrenia PRSs than BD with mood-congruent psychosis and BD without psychosis [12]. On a similar note, higher schizophrenia PRSs associate with psychotic features and earlier age of onset in a sample of BD patients, whereas BD PRSs show a positive correlation with manic symptoms in schizophrenia [15]. Future increases in GWAS sample sizes will provide more accurate effect estimates for SNPs, which in turn can further empower PRSs and their specific relation to psychiatric symptoms across the boundaries of DSM-5 disorder classifications [4].

5.3 Prognosis

The ability to predict the course of a disorder in psychiatric patients could prove highly valuable as for these disorders we do not currently have accurate course prediction models. In psychotic disorder, the potential of PRSs in prediction of conversion from at-risk states to clinical diagnosis has been shown by

Table 1. Clinical utility of genetic cross-disorder analysis in psychiatry.

Clinical domain	Main cross-disorder genetics approaches	Utility	Example
Diagnostics	Genetic correlation, PRS analysis	Describe polygenic overlap between psychiatric disorders	Genetic correlation between schizophrenia and bipolar disorder estimated at 0.80 [3,6,7].
		Describe genetic heterogeneity within one psychiatric disorder	Schizoaffective BD is most strongly associated to schizophrenia polygenic risk, followed by BD-I and BD-II (relative risk based on schizophrenia PRSs of 1.37, 1.30 and 1.04, respectively) [12].
		Describe polygenic overlap with phenotypes outside psychiatry	Schizophrenia genetically correlates with amyotrophic lateral sclerosis (0.14) [8]. MDD (0.11) and ADHD (0.21) have a significant genetic correlation with BMI [7].
	PRS analysis	Identification of clinical subgroups in a heterogeneous disorder (patient stratification)	Schizophrenia PRSs define a gradient in types of psychosis in BD [12].
Prognosis	PRS analysis	Prediction of clinical outcome	First-episode psychosis subjects who later develop schizophrenia have higher schizophrenia PRSs than those who develop other psychotic disorders [16].
		Prediction of predisposition for a disorder	OCD, schizophrenia, schizophrenia-BD and MDD PRSs have predictive value for obsessive compulsive symptoms in a population sample [17]. Future possibility to assess risk for post-trauma psychopathology in groups of healthy military personnel using PRSs.
Treatment	Combined analysis (mega- or meta-analysis)	Detection of shared disease variants that might converge in shared disease pathways	Four pleiotropic loci identified between five psychiatric disorders (ASD, ADHD, BD, MDD and schizophrenia) [11]. 114 shared loci identified between schizophrenia and BD, and four loci with divergent effects [15]. 96 pleiotropy-informed loci identified for depressive symptoms, neuroticism and subjective well-being [19].
		Detection of shared disease variants that are targetable by drugs, or that can induce treatment side-effects	Gene targets of two drugs show enrichment for variants associated with multiple cognitive phenotypes and intellectual disability [20]. Variants associated with clozapine-induced agranulocytosis have been associated with genes involved in adverse response to statins [22].
	PRS analysis	Prediction of treatment response	Schizophrenia PRSs are inversely correlated with lithium response in BD [21].

Abbreviations: PRS, Polygenic Risk Score; ADHD, Attention-Deficit Hyperactivity Disorder; BD, Bipolar Disorder; MDD, Major Depressive Disorder; OCD, Obsessive Compulsive Disorder; ASD, Autism spectrum disorder.

schizophrenia PRSs being significantly higher in individuals with first-episode psychosis later diagnosed with schizophrenia than in first-episode patients diagnosed with other psychotic disorders, whereas both groups had higher PRSs than healthy control subjects [16]. A cross-disorder example includes a population-based sample where PRSs for obsessive compulsive disorder (OCD), schizophrenia, MDD, and combined schizophrenia-BD predicted sub-clinical obsessive compulsive symptoms, thereby showing the potential of these scores to identify individuals who have increased risk to develop OCD and other psychiatric disorders [17]. With regard to psychiatric conditions where environmental factors play a pivotal role, such as post-trauma psychopathology, PRSs for several disorders (such as post-traumatic stress disorder, depression and anxiety) could ideally be used to screen healthy at-risk groups, e.g. military personnel, and take preventative measures for those with high risk scores. However, their currently low accuracy and sensitivity in clinical phenotypes preclude application of such PRSs to risk prediction at the level of the individual. Therefore, integrating PRSs into a diagnostic work-up that includes other data such as psychiatric signs and symptoms may in the near future improve prognostic accuracy [18].

5.4 Treatment

The previously discussed polygenic methods describe the extent to which two disorders correlate, but do not pinpoint specific loci with cross-disorder effects. Information on specific shared disease loci can be obtained by merging individual-level or summary-level data from different disorders in a single association analysis, combining all cases into a single phenotype. Such mega- or meta-analyses increase power to detect pleiotropic loci that were below the genome-wide significance thresholds in single-disorder GWASs. One of the most illustrative cross-disorder studies performed between five psychiatric disorders with the highest heritability estimates (autism spectrum disorder (ASD), ADHD, BD, MDD and schizophrenia) identified four shared genome-wide significant loci with pleiotropic effects which were not genome-wide significant when analyses were performed on disorders separately [11], clearly indicating the increased power in sets of disorders that share genetic loci. Moreover, a combined analysis of schizophrenia and BD individual-level GWAS datasets has identified 114 loci implicating synaptic and neuronal pathways

shared between the two, although these loci exert different effect sizes in both disorders. Interestingly, this study also identified four loci involved with divergent effects between these disorders [15]. Recently, techniques have been developed to perform powerful combined analyses of polygenic traits with summary-level data, e.g. resulting in the discovery of 96 pleiotropy-informed loci in a combined study of depressive symptoms, neuroticism and subjective well-being [19].

Despite the low effect sizes of such pleiotropic variants, they may ultimately prove highly valuable in clinical settings as they inform possible cross-disorder therapeutic targets, empowering drug repurposing. This has for example been shown in a GWAS meta-analysis of phenotypes related to general cognitive ability, where discovered loci were enriched in genes associated with intellectual disability and in gene targets of two pharmacological compounds, prioritizing these substances as possible cognitive enhancers [20]. In addition, PRSs can also be used to predict treatment response, as illustrated by the inverse correlation between schizophrenia PRSs and lithium response in BD patients [21]. Finally, side effects can sometimes be explained by pleiotropic variants, as illustrated by the association of clozapine-induced agranulocytosis with genes previously linked to adverse reactions to statins [22].

5.5 Caveats in genetic cross-disorder studies

Cross-disorder studies should be interpreted with some caution as observations can be driven by other factors than true overlap at causal variants. First, as opposed to a scenario of true biological genetic overlap, a significant genetic correlation could also be observed if a causal variant directly affects disorder A, while disorder B is merely caused by disorder A (mediated pleiotropy). In this scenario, there is no direct effect of the causal variant on disorder B and observed pleiotropy is artificial. Given the differential ages of onset, for psychiatric disorders this would imply that early onset disorders occur first (ASD, ADHD, intellectual disability) and then cause disorders with later onset (MDD, schizophrenia, BD), a notion currently lacking support from the scientific literature. In addition, disorder-specific causal variants in different genes might be tagged by the same SNP due to LD. This SNP could then show an association to both phenotypes while the causal variants are actually different, resulting in

spurious pleiotropy [23]. Second, observed pleiotropy can result from external confounding factors, such as population differences or assortative mating. The latter plays a role in psychiatric disorders, where non-random mating between and within disorders is observed [24]. Third, the abovementioned cross-disorder analysis approaches only use information on SNPs captured in GWAS data, thus omitting other sources of genetic variation, such as rare variants. Fourth, due to overlap in clinical symptoms the risk of diagnostic misclassification is particularly pressing in psychiatry and can lead to false positive findings in cross-disorder studies when patient cohorts are not homogenous. Methodology has been developed to identify heterogeneous subgroups of patients in GWAS datasets (e.g. BUHMBOX) to detect such confounding in cross-disorder analyses [25]. Finally, genotype data of the same healthy controls is often used in multiple GWASs of different disorders. When this overlap is not taken into account in cross-disorder analyses combining these datasets, a false positive correlation – not reflecting shared genetic risk between diseases but merely a correlation induced by genetically identical subjects – may arise. It is therefore imperative to identify and exclude duplicate subjects in cross-disorder studies using individual-level genotype data. Alternatively, various novel cross-disorder methods have been developed that include correction for inflated test results due to sample overlap and allow for the use of only summary-level data [6,19]. Overall, a considerable amount of cross-disorder analyses can nowadays be performed using summary-level data without the requirement of access to individual-level genotype data. As new GWAS results are continuously being published, summary-level data are almost always made publicly available, in line with many journals' manuscript acceptance conditions, whereas individual-level data is often available upon request. Sharing of full GWAS results remains essential for cross-disorder analyses by groups and consortia interested in matching genetic data of other studies with their own datasets.

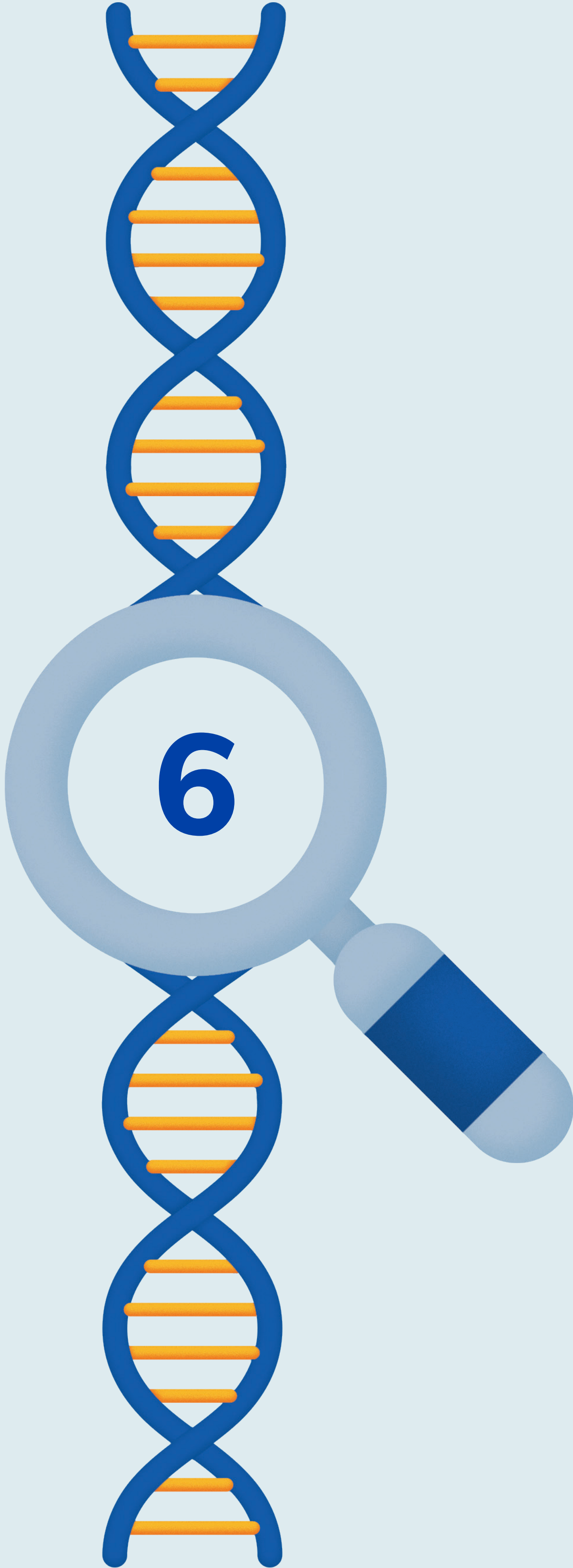
5.6 Conclusions

A range of techniques has been employed to unravel important cross-disorder genetic findings in psychiatry, resulting in intriguing new clinical perspectives at the levels of diagnostics, prognosis and treatment development and prediction.

5.7 References

1. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461(7265):747–53.
2. Nishino J, Kochi Y, Shigemizu D, Kato M, Ikari K, Ochi H, et al. Empirical Bayes Estimation of Semi-parametric Hierarchical Mixture Models for Unbiased Characterization of Polygenic Disease Architectures. Front Genet 2018;9:115.
3. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013;45(9):984–94.
4. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017;101(1):5–22.
5. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 2012;28(19):2540–2.
6. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.
7. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science 2018;360(6395):eaap8757.
8. McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun 2017;8:14774.
9. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 2007;17(10):1520–8.
10. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460(7256):748–52.
11. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013;381(9875):1371–9.
12. Allardyce J, Leonenko G, Hamshere M, Pardiñas AF, Forty L, Knott S, et al. Association Between Schizophrenia-Related Polygenic Liability and the Occurrence and Level of Mood-Incongruent Psychotic Symptoms in Bipolar Disorder. JAMA Psychiatry 2018;75(1):28–35.
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511(7510):421–7.
14. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet 2013;14(7):507–15.
15. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. Cell 2018;173(7):1705–16.
16. Vassos E, Di Forti M, Coleman J, Iyegbe C, Prata D, Euesden J, et al. An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biol Psychiatry 2017;81(6):470–7.
17. Zilhão NR, Abdellaoui A, Smit DJA, Cath DC, Hottenga JJ, Boomsma DI. Polygenic prediction of obsessive compulsive symptoms. Mol Psychiatry 2018;23(2):168–9.
18. Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, et al. Effect of polygenic risk scores on depression in childhood trauma. Br J Psychiatry 2014;205(2):113–9.
19. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet 2018;9:283.
20. Lam M, Trampush JW, Yu J, Knowles E, Davies G, Liewald DC, et al. Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. Cell Rep 2017;21(9):2597–613.

21. International Consortium on Lithium Genetics (ConLi+Gen), Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, et al. Association of Polygenic Score for Schizophrenia and HLA Antigen and Inflammation Genes With Response to Lithium in Bipolar Affective Disorder: A Genome-Wide Association Study. JAMA Psychiatry 2018;75(1):65–74.
22. Legge SE, Hamshere ML, Ripke S, Pardinas AF, Goldstein JI, Rees E, et al. Genome-wide common and rare variant analysis provides novel insights into clozapine-associated neutropenia. Mol Psychiatry 2017;22(10):1502–8.
23. O'Donovan MC, Owen MJ. The implications of the shared genetics of psychiatric disorders. Nat Med 2016;22(11):1214–9.
24. Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. JAMA Psychiatry 2016;73(4):354–61.
25. Han B, Pouget JG, Slowikowski K, Stahl E, Lee CH, Diogo D, et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nat Genet 2016;48(7):803–10.


6

Chapter 6

Genetic correlation between amyotrophic lateral sclerosis and schizophrenia

Dick Schijven*, Russell L. McLaughlin*, Wouter van Rheenen, Kristel R. van Eijk, Margaret O'Brien, Project MinE ALS GWAS Consortium#, Schizophrenia Working Group of the Psychiatric Genomics Consortium#, René S. Kahn, Roel A. Ophoff, An Goris, Daniel G. Bradley, Ammar Al-Chalabi, Leonard H. van den Berg, Jurjen J. Luykx**, Orla Hardiman**, Jan H. Veldink**

*Equal contributions | **Joint supervision | #Consortium members are listed in appendix I

Published in *Nature Communications*, 8:14774 (2017)

Abstract

We have previously shown higher-than-expected rates of schizophrenia in relatives of patients with amyotrophic lateral sclerosis (ALS), suggesting an etiological relationship between the diseases. Here, we investigate the genetic relationship between ALS and schizophrenia using genome-wide association study data from over 100,000 unique individuals. Using linkage disequilibrium score regression, we estimate the genetic correlation between ALS and schizophrenia to be 14.3% (7.05–21.6; $p = 1 \times 10^{-4}$) with schizophrenia polygenic risk scores explaining up to 0.12% of the variance in ALS ($p = 8.4 \times 10^{-7}$). A modest increase in comorbidity of ALS and schizophrenia is expected given these findings (odds ratio 1.08–1.26) but this would require very large studies to observe epidemiologically. We identify five potential novel ALS-associated loci using conditional false discovery rate analysis. It is likely that shared neurobiological mechanisms between these two disorders will engender novel hypotheses in future preclinical and clinical studies.

6.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a late-onset neurodegenerative condition characterized by progressive loss of upper and lower motor neurons, leading to death from respiratory failure in 70% of patients within 3 years of symptom onset. Although ALS is often described as a primarily motor-system disease, extramotor involvement occurs in up to 50% of cases, with prominent executive and behavioral impairment, and behavioral variant frontotemporal dementia (FTD) in up to 14% of cases [1]. A neuropsychiatric prodrome has been described in some people with ALS-FTD, and higher rates of schizophrenia and suicide have been reported in first and second-degree relatives of those with ALS, particularly in kindreds associated with the *C9orf72* hexanucleotide repeat expansion [2]. These clinical and epidemiological observations suggest that ALS and schizophrenia may share heritability.

ALS and schizophrenia both have high heritability estimates (0.65 and 0.64, respectively) [3,4]; however, the underlying genetic architectures of these heritable components appear to differ. Analysis of large genome-wide association study (GWAS) datasets has implicated over 100 independent risk loci for schizophrenia [5] and estimated that a substantial proportion (23%) of the variance in underlying liability for schizophrenia is due to additive polygenic risk (many risk-increasing alleles of low individual effect combining to cause disease) conferred by common genetic variants [6]. This proportion, the single nucleotide polymorphism (SNP)-based heritability, is lower in ALS (8.2%), in which fewer than ten risk loci have been identified by GWAS [7]. Nevertheless, both diseases have polygenic components, but the extent to which they overlap has not been investigated.

Recently, methods to investigate overlap between polygenic traits using GWAS data have been developed [8-10]. These methods assess either pleiotropy (identical genetic variants influencing both traits) or genetic correlation (identical alleles influencing both traits). Genetic correlation is related to heritability; for both measures, binary traits such as ALS and schizophrenia are typically modelled as extremes of an underlying continuous scale of liability to develop the trait. If two binary traits are genetically correlated, their liabilities covary, and this covariance is determined by both traits having identical risk

alleles at overlapping risk loci. Studies of pleiotropy and genetic correlation have provided insights into the overlapping genetics of numerous traits and disorders, although none to date has implicated shared polygenic risk between neurodegenerative and neuropsychiatric disease. Here, we apply several techniques to identify and dissect the polygenic overlap between ALS and schizophrenia. We provide evidence for genetic correlation between the two disorders which is unlikely to be driven by diagnostic misclassification and we demonstrate a lack of polygenic overlap between ALS and other neuropsychiatric and neurological conditions, which could be due to limited power given the smaller cohort sizes for these studies.

6.2 Results

6.2.1 Genetic correlation between ALS and schizophrenia

To investigate the polygenic overlap between ALS and schizophrenia, we used individual-level and summary data from GWAS for ALS [7] (36,052 individuals) and schizophrenia [5] (79,845 individuals). At least 5,582 control individuals were common to both datasets, but for some cohorts included in the schizophrenia dataset this could not be ascertained so this number is likely to be higher. For ALS, we used summary data from both mixed linear model association testing [11] and meta-analysis of cohort-level logistic regression [12]. We first used linkage disequilibrium (LD) score regression with ALS and schizophrenia summary statistics; this technique models, for polygenic traits, a linear relationship between a SNP's LD score (the amount of genetic variation that it captures) and its GWAS test statistic [13]. This distinguishes confounding from polygenicity in GWAS inflation and the regression coefficient can be used to estimate the SNP-based heritability (h_S^2) for single traits [13]. In the bivariate case, the regression coefficient estimates genetic covariance (ρ_g) for pairs of traits, from which genetic correlation (r_g) is estimated [8]; these estimates are unaffected by sample overlap between traits. Using constrained intercept LD score regression with mixed linear model ALS summary statistics, we estimated the liability-scale SNP-based heritability of ALS to be 8.2% (95% confidence interval = 7.2-9.1; mean χ^2 = 1.13; all ranges reported below indicate 95% confidence intervals), replicating previous estimates based on alternative methods [7]. Estimates based on ALS meta-analysis summary statistics and free-intercept LD score regression with mixed linear model summary statistics

were lower (Supplementary Table 1), resulting in higher genetic correlation estimates (Supplementary Table 2); for this reason, we conservatively use constrained intercept genetic correlation estimates for ALS mixed linear model summary statistics throughout the remainder of this paper. Heritability estimates for permuted ALS data were null (Supplementary Table 1).

LD score regression estimated the genetic correlation between ALS and schizophrenia to be 14.3% (7.05-21.6; $p = 1 \times 10^{-4}$). Results were similar for a smaller schizophrenia cohort of European ancestry (21,856 individuals) [14], indicating that the inclusion of individuals of Asian ancestry in the schizophrenia cohort did not bias this result (Supplementary Figure 1). In addition to

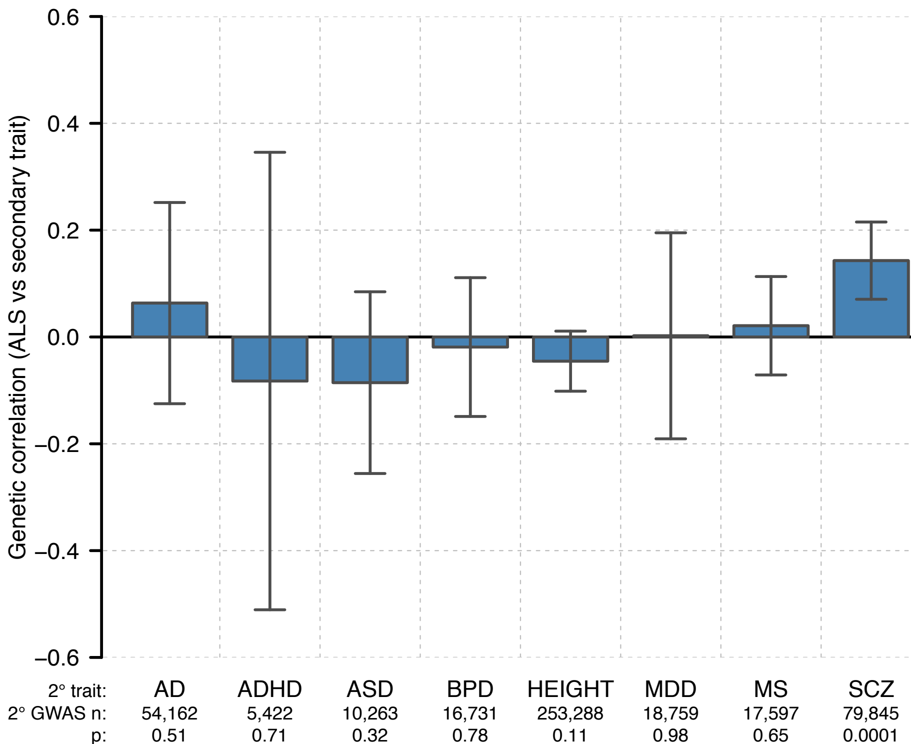


Figure 1. Genetic correlation between ALS and eight secondary traits. Error bars indicating 95% confidence intervals and p-values were calculated by the LD score regression software using a block jackknife procedure. Secondary traits are: AD, Alzheimer's disease; ADHD, attention deficit-hyperactivity disorder; ASD, autism spectrum disorder; BPD, bipolar disorder; MDD, major depressive disorder; MS, multiple sclerosis; SCZ, schizophrenia.

schizophrenia, we estimated genetic correlation with ALS using GWAS summary statistics for bipolar disorder [15], major depressive disorder [16], attention deficit-hyperactivity disorder [17], autism spectrum disorder [17], Alzheimer's disease (Supplementary note 1) [18], multiple sclerosis [19] and adult height [20], finding no significant genetic correlation between ALS and any secondary trait other than schizophrenia (Figure 1; Supplementary Table 2).

6.2.2 Polygenic risk score analysis

We supported the positive genetic correlation between ALS and schizophrenia by analysis of polygenic risk for schizophrenia in the ALS cohort. Polygenic risk scores (PRS) are per-individual scores based on the sum of alleles associated

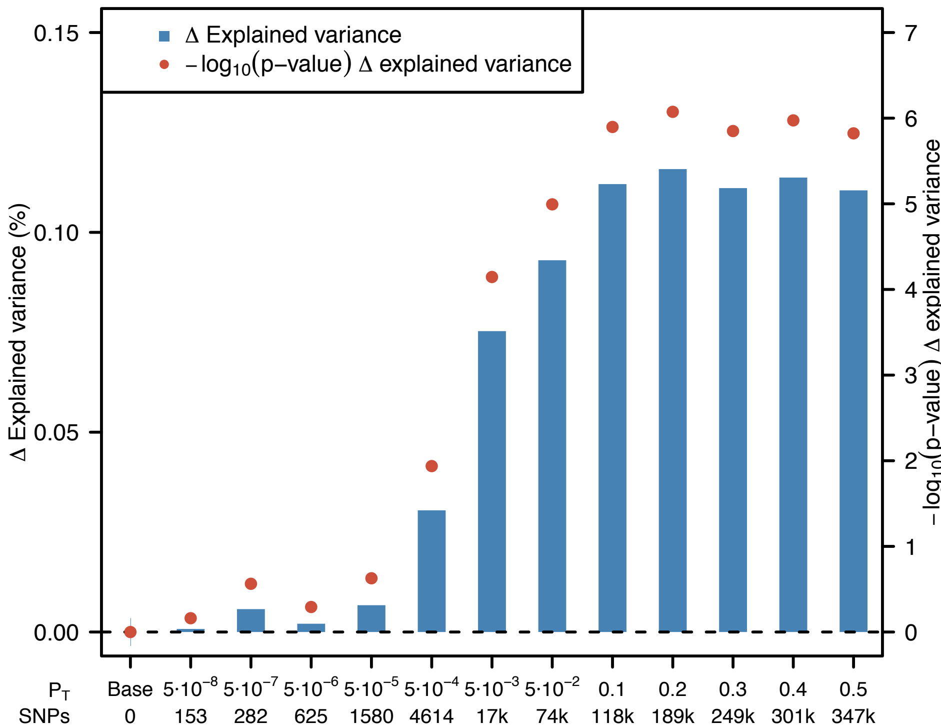


Figure 2. Analysis of PRS for schizophrenia in a target sample of 10,032 ALS cases and 16,627 healthy controls. P-value thresholds (P_T) for schizophrenia SNPs are shown on the x-axis, where the number of SNPs increases with a more lenient P_T. Δ Explained variances (Nagelkerke R^2, shown as a %) of a generalized linear model including schizophrenia-based PRS versus a baseline model without polygenic scores (blue bars) are shown for each P_T. $-\text{Log}_{10}$ p-values of Δ explained variance per P_T (red dots) represent p-values from the binomial logistic regression of ALS phenotype on PRS, accounting for LD (Supplementary Table 4) and including sex and significant principal components as covariates (Supplementary Figure 2). Values are provided in Supplementary Table 5.

with one phenotype, weighted by their effect size, measured in an independent target sample of the same or a different phenotype [10]. PRS calculated on schizophrenia GWAS summary statistics for twelve P-value thresholds (P_T) explained up to 0.12% (P_T = 0.2, p = 8.4 x 10^{-7}) of the phenotypic variance in a subset of the individual-level ALS genotype data that had all individuals removed that were known or suspected to be present in the schizophrenia cohort (Figure 2; Supplementary Table 5).

ALS cases had on average higher PRS for schizophrenia compared to healthy controls and harboring a high schizophrenia PRS for P_T = 0.2 significantly increased the odds of being an ALS patient in our cohort (Figure 3;

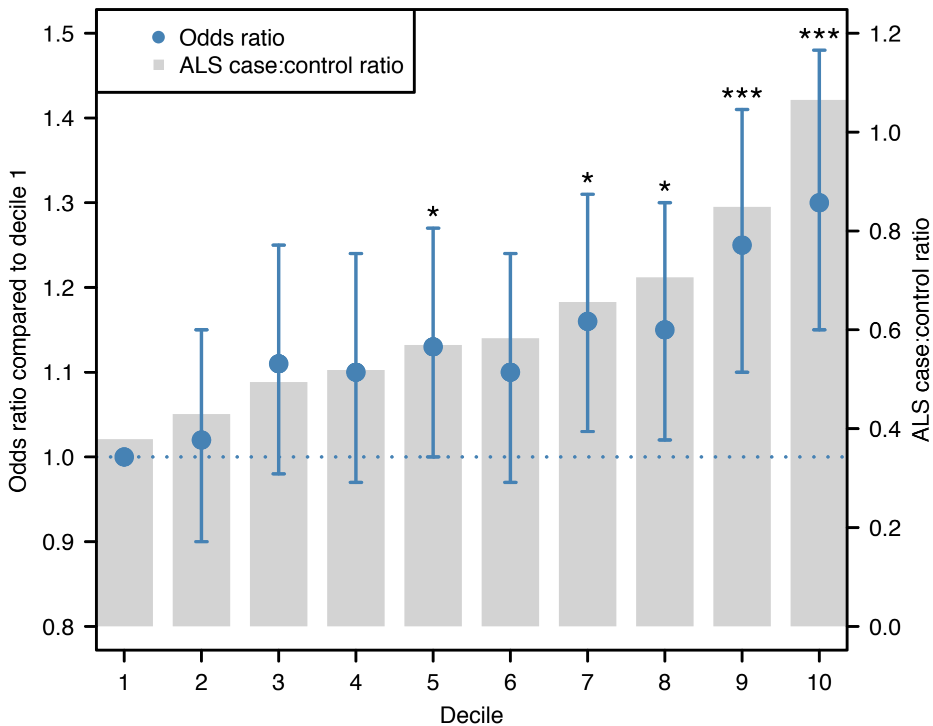


Figure 3. Odds ratio for ALS by PRS deciles for schizophrenia. The figure applies to schizophrenia p-value threshold (P_T)=0.2. The PRS for this threshold were converted to ten deciles containing near identical numbers of individuals. Decile 1 contained the lowest scores and decile 10 contained the highest scores, where decile 1 was the reference and deciles 2–10 were dummy variables to contrast to decile 1 for OR calculation. The case:control ratio per decile is indicated with grey bars. Error bars indicate 95% confidence intervals. Significant differences from decile 1 were determined by logistic regression of ALS phenotype on PRS decile, including sex and principal components as covariates and are indicated by * p < 0.05 or *** p < 0.001.

Supplementary Table 6). Permutation of case–control labels reduced the explained variance to values near zero (Supplementary Figure 3).

6.2.3 Modelling misdiagnosis and comorbidity

Using BUHMBOX [21], a tool that distinguishes true genetic relationships between diseases (pleiotropy) from spurious relationships resulting from heterogeneous mixing of disease cohorts, we determined that misdiagnosed cases in the schizophrenia cohort (for example, young-onset FTD–ALS) did not drive the genetic correlation estimate between ALS and schizophrenia (p = 0.94). Assuming a true genetic correlation of 0%, we estimated the required rate of misdiagnosis of ALS as schizophrenia to be 4.86% (2.47-7.13) to obtain the genetic correlation estimate of 14.3% (7.05-21.6; Supplementary Table 7), which we consider to be too high to be likely. However, if ALS and schizophrenia are genetically correlated, more comorbidity would be expected than if the genetic correlation was 0%. Modelling our observed genetic correlation of 14.3% (7.05-21.6), we estimated the odds ratio for having above-threshold liability for ALS given above-threshold liability for schizophrenia to be 1.17 (1.08-1.26), and the same for schizophrenia given ALS (Supplementary Figure 4). From a clinical perspective, to achieve 80% power to detect a significant

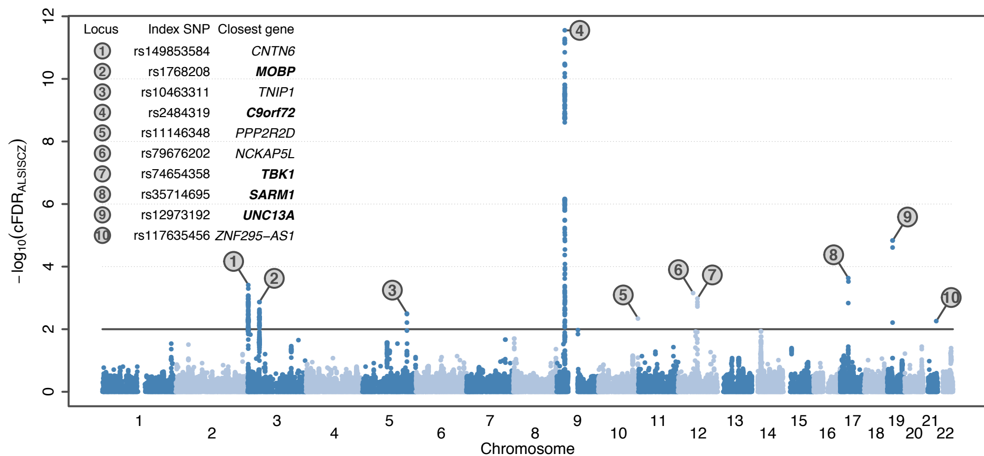


Figure 4. Pleiotropy-informed ALS risk loci determined by analysis of cFDR in ALS GWAS p-values given schizophrenia GWAS p-values ($cFDR_{ALS|SCZ}$). Each point denotes a SNP; its x-axis position corresponds to its chromosomal location and its height indicates the extent of association with ALS by cFDR analysis. The solid line indicates the threshold cFDR = 0.01. Any gene whose role in ALS is already established is in bold. A complete list of all loci at cFDR ≤ 0.05 is provided in Supplementary Table 8.

($\alpha = 0.05$) excess of schizophrenia in the ALS cohort as a result of this genetic correlation, the required population-based incident cohort size is 16,448 ALS patients (7,310–66,670).

6.2.4 Pleiotropic risk loci

We leveraged the genetic correlation between ALS and schizophrenia to discover novel ALS-associated genomic loci by conditional false discovery rate (cFDR) analysis [9,22] (Figure 4; Supplementary Table 8). Five loci already known to be involved in ALS were identified (corresponding to *MOBP*, *C9orf72*, *TBK1*, *SARM1* and *UNC13A*) along with five potential novel loci at cFDR < 0.01 (*CNTN6*, *TNIP1*, *PPP2R2D*, *NCKAP5L* and *ZNF295-AS1*). No gene set was significantly enriched (after Bonferroni correction) in genome-wide cFDR values when analyzed using MAGENTA.

6.3 Discussion

There is evolving clinical, epidemiological and biological evidence for an association between ALS and psychotic illness, particularly schizophrenia. Genetic evidence of overlap to date has been based primarily on individual genes showing Mendelian inheritance, in particular the *C9orf72* hexanucleotide repeat expansion, which is associated with ALS and FTD, and with psychosis in relatives of ALS patients [2]. In this study, we have replicated SNP-based heritability estimates for ALS and schizophrenia using GWAS summary statistics, and have for the first time demonstrated significant overlap between the polygenic components of both diseases, estimating the genetic correlation to be 14.3%. We have carefully controlled for confounding bias, including population stratification and shared control samples, and have shown through analysis of polygenic risk scores that the overlapping polygenic risk applies to SNPs that are modestly associated with both diseases. Given that our genetic correlation estimate relates to the polygenic components of ALS ($h_S^2 = 8.2\%$) and schizophrenia ($h_S^2 = 23\%$) and these estimates do not represent all heritability for both diseases, the accuracy of using schizophrenia-based PRS to predict ALS status in any patient is expected to be low (Nagelkerke's $R^2 = 0.12\%$ for $P_T = 0.2$), although statistically significant ($p = 8.4 \times 10^{-7}$). Nevertheless, the positive genetic correlation of 14.3% indicates that the direction of effect of risk-increasing and protective alleles is consistently

aligned between ALS and schizophrenia, suggesting convergent biological mechanisms between the two diseases.

Although phenotypically heterogeneous, both ALS and schizophrenia are clinically recognizable as syndromes [23,24]. The common biological mechanisms underlying the association between the two conditions are not well understood, but are likely associated with disruption of cortical networks. Schizophrenia is a polygenic neurodevelopmental disorder characterized by a combination of positive symptoms (hallucinations and delusions), negative symptoms (diminished motivation, blunted affect, reduction in spontaneous speech and poor social functioning) and impairment over a broad range of cognitive abilities [25]. ALS is a late onset complex genetic disease characterized by a predominantly motor phenotype with recently recognized extra-motor features in 50% of patients, including cognitive impairment [1]. It has been suggested that the functional effects of risk genes in schizophrenia converge by modulating synaptic plasticity, and influencing the development and stabilization of cortical microcircuitry [5]. In this context, our identification of *CNTN6* (contactin 6, also known as NB-3, a neural adhesion protein important in axon development) [26] as a novel pleiotropy-informed ALS-associated locus supports neural network dysregulation as a potential convergent mechanism of disease in ALS and schizophrenia.

No significantly enriched biological pathway or ontological term was identified within genome-wide cFDR values using MAGENTA. Low inflation in ALS GWAS statistics, coupled with a rare variant genetic architecture [7], render enrichment-based biological pathway analyses with current sample sizes challenging. Nevertheless, nine further loci were associated with ALS risk at cFDR < 0.01. Of these, *MOBP*, *C9orf72*, *TBK1*, *SARM1* and *UNC13A* have been described previously in ALS and were associated by cFDR analysis in this study owing to their strong association with ALS through GWAS [7]. The remaining four loci (*TNIP1*, *PPP2R2D*, *NCKAP5L* and *ZNF295-AS1*) are novel associations and may represent pleiotropic disease loci. *TNIP1* encodes TNFAIP3 interacting protein 1 and is involved in autoimmunity and tissue homoeostasis [27]. The protein product of *PPP2R2D* is a regulatory subunit of protein phosphatase 2 and has a role in PI3K-Akt signaling and mitosis [28]. *NCKAP5L* is a homologue of *NCKAP5*, encoding NAP5, a proline-rich protein that has previously been

implicated in schizophrenia, bipolar disorder and autism [29,30]. *ZNF295-AS1* is a noncoding RNA [31]. Further investigation into the biological roles of these genes may yield novel insight into the pathophysiology of certain subtypes of ALS and schizophrenia, and as whole-genome and exome datasets become available in the future for appropriately large ALS case-control cohorts, testing for burden of rare genetic variation across these genes will be particularly instructive, especially given the role that rare variants appear to play in the pathophysiology of ALS [7].

Our data suggest that other neuropsychiatric conditions (bipolar disorder, autism and major depression) do not share polygenic risk with ALS. This finding contrasts with our recent observations from family aggregation studies and may be unexpected given the extensive genetic correlation between neuropsychiatric conditions [6]. This could relate to statistical power conferred by secondary phenotype cohort sizes, and future studies with larger sample sizes will shed further light on the relationship between ALS and neuropsychiatric disease. It is also possible that the current study underestimates genetic correlations due to the substantial role that rare variants play in the genetic architecture of ALS [7] and future fine-grained studies examining heritability and genetic correlation in low-minor allele frequency and low-LD regions may identify a broader relationship between ALS and neuropsychiatric diseases.

A potential criticism of this study is that the polygenic overlap between ALS and schizophrenia could be driven by misdiagnosis, particularly in cases of ALS–FTD, which can present in later life as a psychotic illness and could be misdiagnosed as schizophrenia. This is unlikely, as strict diagnostic criteria are required for inclusion of samples in the schizophrenia GWAS dataset [5]. Furthermore, since core schizophrenia symptoms are usually diagnosed during late adolescence, a misdiagnosis of FTD-onset ALS-FTD as schizophrenia is unlikely. In this study, we found no evidence for misdiagnosis of ALS as schizophrenia (BUHMBOX $p = 0.94$) and we estimated that a misdiagnosis of 4.86% of ALS cases would be required to spuriously observe a genetic correlation of 14.3%, which is not likely to occur in clinical practice. We are therefore confident that this genetic correlation estimate reflects a genuine polygenic overlap between the two diseases and is not a feature of cohort

ascertainment, but the possibility of some misdiagnosis in either cohort cannot be entirely excluded based on available data.

A positive genetic correlation between ALS and schizophrenia predicts an excess of patients presenting with both diseases. Most neurologists and psychiatrists, however, will not readily acknowledge that these conditions co-occur frequently. Our genetic correlation estimate confers an odds ratio of 1.17 (1.08-1.26) for harboring above-threshold liability for ALS given schizophrenia (or vice versa) and a lifetime risk of 1:34,336 for both phenotypes together. Thus, a very large incident cohort of 16,448 ALS patients (7,310-66,670), with detailed phenotype information, would be required to have sufficient power to detect an excess of schizophrenia within an ALS cohort. Coupled with reduced life expectancy in patients with schizophrenia [32], this may explain the relative dearth of epidemiological studies to date providing clinical evidence of excess comorbidity. Moreover, it has also been proposed that prolonged use of antipsychotic medication may protect against developing all of the clinical features of ALS [33], which would reduce the rate of observed comorbidity. Considering our novel evidence for a genetic relationship between ALS and schizophrenia, this underscores the intriguing possibility that therapeutic strategies for each condition may be useful in the other, and our findings provide rationale to consider the biology of ALS and schizophrenia as related in future drug development studies. Indeed, the glutamate-modulating ALS therapy riluzole has shown efficacy as an adjunct to risperidone, an antipsychotic medication, in reducing the negative symptoms of schizophrenia [34].

In conclusion, we have estimated the genetic correlation between ALS and schizophrenia to be 14.3% (7.05-21.6), providing molecular genetic support for our epidemiological observation of psychiatric endophenotypes within ALS kindreds. To our knowledge, this is the first study to show genetic correlation derived from polygenic overlap between neurodegenerative and neuropsychiatric phenotypes. The presence of both apparent monogenic *C9orf72*-driven overlap [2] and polygenic overlap in the etiology of ALS and schizophrenia suggests the presence of common biological processes, which may relate to disruption of cortical circuitry. As both ALS and schizophrenia are heterogeneous conditions, further genomic, biological and clinical studies are likely to yield novel insights into the pathological processes for both diseases

and will provide clinical sub-stratification parameters that could drive novel drug development for both neurodegenerative and psychiatric conditions.

6.4 Methods

6.4.1 Study population and genetic data

For ALS, 7,740,343 SNPs genotyped in 12,577 ALS patients and 23,475 healthy controls of European ancestry organized in 27 platform- and country-defined strata were used [7]. The schizophrenia dataset comprised GWAS summary statistics for 9,444,230 SNPs originally genotyped in 34,241 patients and 45,604 controls of European and Asian ancestry [5]. For LD score regression, GWAS summary statistics were generated for the ALS cohort using mixed linear model association testing implemented in Genome-wide Complex Trait Analysis [11] or logistic regression combined with cross-stratum meta-analysis using METAL [12]. To evaluate sample overlap for PRS and cFDR analyses, we also obtained individual-level genotype data for 27,647 schizophrenia cases and 33,675 controls from the schizophrenia GWAS (Psychiatric Genomics Consortium [5] and dbGaP accession number phs000021.v3.p2). Using 88,971 LD-pruned (window size 200 SNPs; shift 20 SNPs; $R^2 > 0.25$) SNPs in both datasets (INFO score > 0.8; MAF > 0.2), with SNPs in high-LD regions removed (Supplementary Table 4), samples were removed from the ALS dataset if they were duplicated or had a cryptically related counterpart (PLINK PIHAT > 0.1; 5,582 individuals) in the schizophrenia cohort and whole strata (representing Finnish and German samples; 3,811 individuals) were also removed if commonality with the schizophrenia cohort could not be ascertained (due to unavailability of individual-level genotype data in the schizophrenia cohort) and in which a sample overlap was suspected (Supplementary Table 3).

6.4.2 LD score regression

We calculated LD scores using LDSC v1.0.0 in 1 centiMorgan windows around 13,307,412 non-singleton variants genotyped in 379 European individuals (CEU, FIN, GBR, IBS and TSI populations) in the phase 1 integrated release of the 1,000 Genomes Project [35]. For regression weights [13], we restricted LD score calculation to SNPs included in both the GWAS summary statistics and HapMap phase 3; for r_g estimation in pairs of traits this was the intersection of SNPs for both traits and HapMap. Because population structure and

6

confounding were highly controlled in the ALS summary statistics by the use of mixed linear model association tests, we constrained the LD score regression intercept to 1 for h_S^2 estimation in ALS, and we also estimated h_S^2 with a free intercept. For h_S^2 estimation in all other traits and for r_g estimation the intercept was a free parameter. We also estimated r_g using ALS meta-analysis results [7] with free and constrained intercepts and with permuted data conserving population structure. Briefly, principal component analysis was carried out for each stratum using smartpca [36] and the three-dimensional space defined by principal components 1-3 was equally subdivided into 1,000 cubes. Within each cube, case-control labels were randomly swapped and association statistics were re-calculated for the entire stratum using logistic regression. Study-level p-values were then calculated using inverse variance weighted fixed effect meta-analysis implemented in METAL [7,12]. h_S^2 was estimated for these meta-analyzed permuted data using LD score regression (Supplementary Table 1).

6.4.3 Polygenic risk score analysis

We calculated PRS for 10,032 cases and 16,627 healthy controls in the ALS dataset (duplicate and suspected or confirmed related samples with the schizophrenia dataset removed), based on schizophrenia-associated alleles and effect sizes reported in the GWAS summary statistics for 6,843,674 SNPs included in both studies and in the phase 1 integrated release of the 1,000 Genomes Project [35] (imputation INFO score < 0.3; minor allele frequency < 0.01; A/T and G/C SNPs removed). SNPs were clumped in two rounds (physical distance threshold of 250 kb and a LD threshold (R^2) of > 0.5 in the first round and a distance of 5,000 kb and LD threshold of > 0.2 in the second round) using PLINK v1.90b3y, removing high-LD regions (Supplementary Table 4), resulting in a final set of 496,548 SNPs for PRS calculations. Odds ratios for autosomal SNPs reported in the schizophrenia summary statistics were log-converted to beta values and PRS were calculated using PLINK's score function for twelve schizophrenia GWAS P-value thresholds (P_T): 5×10^{-8}, 5×10^{-7}, 5×10^{-6}, 5×10^{-5}, 5×10^{-4}, 5×10^{-3}, 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5. A total of 100 principal components (PCs) were generated for the ALS sample using GCTA version 1.24.4. Using R version 3.2.2, a generalized linear model was applied to model the phenotype of individuals in the ALS dataset. PCs that had a significant effect on the phenotype ($p < 0.0005$, Bonferroni-corrected for 100 PCs) were

selected (PCs 1, 4, 5, 7, 8, 10, 11, 12, 14, 36, 49). To estimate explained variance of PRS on the phenotype, a baseline linear relationship including only sex and significant PCs as variables was modelled first:

$$y = \alpha + \beta_{\mathrm{sex}} x_{\mathrm{sex}} + \sum_{\mathrm{n}} \beta_{\mathrm{pc_n}} x_{\mathrm{pc_n}} \quad \text{(Equation 1)}$$

where y is the phenotype in the ALS dataset, α is the intercept of the model with a slope β for each variable x. Subsequently, a linear model including polygenic scores for each schizophrenia P_T was calculated:

$$y = \alpha + \beta_{\mathrm{sex}} x_{\mathrm{sex}} + \sum_{\mathrm{n}} \beta_{\mathrm{pc_n}} x_{\mathrm{pc_n}} + \beta_{\mathrm{prs}} x_{\mathrm{prs}} \quad \text{(Equation 2)}$$

A Nagelkerke R^2 value was obtained for every model and the baseline Nagelkerke R^2 value was subtracted, resulting in a Δ explained variance that describes the contribution of schizophrenia-based PRS to the phenotype in the ALS dataset. PRS analysis was also performed in permuted case–control data (1,000 permutations, conserving case–control ratio) to assess whether the increased Δ explained variance was a true signal associated with phenotype. Δ explained variances and p-values were averaged across permutation analyses. To ensure we did not over- or under-correct for population effects in our model, we tested the inclusion of up to a total of 30 PCs in the model, starting with the PC with the most significant effect on the ALS phenotype (Supplementary Figure 2). Increasing the number of PCs initially had a large effect on the Δ explained variance, but this effect levelled out after 11 PCs. On the basis of this test we are confident that adding the 11 PCs that had a significant effect on the phenotype sufficiently accounted for possible confounding due to population differences.

For the schizophrenia P_T for which we obtained the highest Δ explained variance (0.2), we subdivided observed schizophrenia-based PRS in the ALS cohort into deciles and calculated the odds ratio for being an ALS case in each decile compared to the first decile using a similar generalized linear model:

$$y = \alpha + \beta_{\mathrm{sex}} x_{\mathrm{sex}} + \sum_{\mathrm{n}} \beta_{\mathrm{pc_n}} x_{\mathrm{pc_n}} + \beta_{\mathrm{decile}} x_{\mathrm{decile}} \quad \text{(Equation 3)}$$

Odds ratios and 95% confidence intervals for ALS were derived by calculating the exponential function of the beta estimate of the model for each of the deciles 2–10.

6

6.4.4 Diagnostic misclassification

To distinguish the contribution of misdiagnosis from true genetic pleiotropy we used BUHMBOX [21] with 417 independent ALS risk alleles in a sample of 27,647 schizophrenia patients for which individual-level genotype data were available. We also estimated the required misdiagnosis rate *M* of FTD-ALS as schizophrenia that would lead to the observed genetic correlation estimate as $C / (C + 1)$, where $C = \rho_g N_{SCZ} / N_{ALS}$ and N_{SCZ} and N_{ALS} are the number of cases in the schizophrenia and ALS datasets, respectively [37] (derived in Supplementary Methods 1).

6.4.5 Expected comorbidity

To investigate the expected comorbidity of ALS and schizophrenia given the observed genetic correlation, we modelled the distribution in liability for ALS and schizophrenia as a bivariate normal distribution with the liability-scale covariance determined by LD score regression (Supplementary Methods 2). Lifetime risks for ALS [38] and schizophrenia [25] of 1/400 and 1/100, respectively, were used to calculate liability thresholds above which individuals develop ALS or schizophrenia, or both. The expected proportions of individuals above these thresholds were used to calculate the odds ratio of developing ALS given schizophrenia, or vice versa (Supplementary Methods 2). The required population size to observe a significant excess of comorbidity was calculated using the binomial power equation.

6.4.6 Pleiotropy-informed risk loci for ALS

Using an adapted cFDR method [39] that allows shared controls between cohorts [22], we estimated per-SNP cFDR given LD score-corrected [8] schizophrenia GWAS P-values for ALS mixed linear model summary statistics calculated in a dataset excluding Finnish and German cohorts (in which suspected control overlap could not be determined), but including all other overlapping samples (totaling 5,582). To correct for the relationship between LD and GWAS test statistics, schizophrenia summary statistics were residualized on LD score by subtracting the product of each SNP's LD score and the univariate LD score regression coefficient for schizophrenia. cFDR values conditioned on these residualized schizophrenia GWAS P-values were calculated for mixed linear model association statistics calculated at 6,843,670 SNPs genotyped in 10,147 ALS cases and 22,094 controls. Pleiotropic genomic

loci were considered statistically significant if cFDR < 0.01 (following Andreassen *et al.* [39]) and were clumped with all neighboring SNPs based on LD ($R^2 > 0.1$) in the complete ALS dataset. Associated cFDR genomic regions were then mapped to the locations of known RefSeq transcripts in human genome build GRCh37. Genome-wide cFDR values were also tested for enrichment in 9,711 gene sets included in the MAGENTA software package (version 2.4, July 2011) and derived from databases such as Gene Ontology (GO, www.geneontology.org), Kyoto Encyclopedia of Genes and Genomes (KEGG, www.kegg.jp), Protein ANalysis THrough Evolutionary Relationships (PANTHER, www.pantherdb.org) and INGENUITY (www.ingenuity.com). SNPs were mapped to genes including 20 kb up- and downstream regions to include regulatory elements. The enrichment cutoff applied in our analysis was based on the 95th percentile of gene scores for all genes in the genome. The null distribution of gene scores for each gene set was based on 10,000 randomly sampled gene sets with equal size. MAGENTA uses a Mann-Whitney rank-sum test to assess gene-set enrichment [40].

6.5 Supplementary information

6.5.1 Supplementary methods 1: Estimation of diagnostic misclassification

A positive genetic correlation between ALS and schizophrenia could derive from misdiagnosis of cognitive- or behavioral-onset FTD-ALS as schizophrenia. For traits A and B, Wray *et al.* [37] describe the relationship between the misdiagnosis rate of trait A as trait B (M_{TA}) and the number of cases diagnosed with trait A (N_{caseDA}), the genetic variance of trait A ($\sigma_{uDA}{}^2$) and the genetic covariance of traits A and B ($\sigma_{uDA,UDB}$) as:

$$N_{\text{caseDA}} = (1 - M_{\text{TA}})N_{\text{caseTA}} + M_{\text{TB}}N_{\text{caseTB}}\,, \qquad \text{(Equation 4)}$$

$$\sigma^2_{\text{uDA}} = \frac{(1 - M_{\text{TA}})^2 N^2_{\text{caseTA}}\sigma^2_{\text{uTA}} + M^2_{\text{TB}}N^2_{\text{caseTB}}\sigma^2_{\text{uTB}} + 2(1 - M_{\text{TA}})M_{\text{TB}}N_{\text{caseTA}}N_{\text{caseTB}}\sigma_{\text{uTA,uTB}}}{N^2_{\text{caseDA}}}\,, \qquad \text{(Eq. 5)}$$

and

$$\sigma_{\text{uDA,uDB}} = \frac{[\left((1 - M_{\text{TA}})(1 - M_{\text{TB}}) + M_{\text{TA}}M_{\text{TB}}\right)N_{\text{caseTA}}N_{\text{caseTB}}\sigma_{\text{uTA,uTB}} + (1 - M_{\text{TA}})M_{\text{TA}}N^2_{\text{caseTA}} + (1 - M_{\text{TB}})M_{\text{TB}}N^2_{\text{caseTB}}\sigma^2_{\text{uTB}}]}{N_{\text{caseDA}}N_{\text{caseDB}}}\,, \text{(Eq. 6)}$$

Where N_{caseTA} and N_{caseTB} are the total numbers of “true” cases of traits A and B respectively, M_{TB} is the misdiagnosis rate of trait B as trait A, $\sigma_{uTA}{}^2$ and and $\sigma_{uTB}{}^2$

are the genetic variances of true cases of traits A and B respectively, $\sigma_{uTA,uTB}$ is the genetic covariance of true cases of traits A and B and N_{caseDB} is the number of cases diagnosed with trait B. For consistency with Wray *et al.* [37] we have used their symbols in these derivations; in the main text of this study the symbol M is used for M_{TA}, ρ_g for $\sigma_{uDA,uDB}$, N_{ALS} for N_{caseDA} and N_{SCZ} for N_{caseDB}.
Assigning ALS as trait A and schizophrenia as trait B, and reasoning that misdiagnosis of schizophrenia as ALS is highly unlikely ($M_{TB} = 0$), our specific interest is in cases where FTD-ALS (trait A) has been misdiagnosed as schizophrenia (trait B) under a true genetic correlation of 0 (ie $\sigma_{uTA,uTB} = 0$). Equations 1-3 therefore simplify to:

$$N_{caseDA} = (1 - M_{TA})N_{caseTA} \,, \qquad \text{(Equation 7)}$$

$$\sigma^2_{uDA} = \frac{(1 - M_{TA})^2 N^2_{caseTA} \sigma^2_{uTA}}{N^2_{caseDA}} \qquad \text{(Equation 8)}$$

and

$$\sigma_{uDA,uDB} = \frac{(1 - M_{TA}) M_{TA} N^2_{caseTA}}{N_{caseDA} N_{caseDB}} \,. \qquad \text{(Equation 9)}$$

The assumption $M_{TB} = 0$ implies that the genetic variance estimate of diagnosed cases of trait A ($\sigma_{DA}{}^2$) accurately reflects the genetic variance of true cases of trait A ($\sigma_{TA}{}^2$); this is confirmed by substituting equation 7 into equation 8:

$$\sigma^2_{DA} = \frac{(1 - M_{TA})^2 N^2_{caseTA} \sigma^2_{uTA}}{(1 - M_{TA})^2 N^2_{caseTA}} \,,$$

which simplifies to $\sigma_{uDA}{}^2 = \sigma_{uTA}{}^2$. The misdiagnosis rate of trait A (M_{TA}) can be established by rearranging equation 7 in terms of the unknown term N_{caseTA}:

$$N_{caseTA} = \frac{N_{caseDA}}{(1 - M_{TA})} \,,$$

and substituting this into equation 9 to give the relationship between genetic covariance ($\sigma_{uDA,uDB}$) and M_{TA}, N_{caseDA} and N_{caseDB}:

$$\sigma_{uDA,uDB} = \frac{(1 - M_{TA}) M_{TA} \dfrac{N^2_{caseDA}}{(1 - M_{TA})^2}}{N_{caseDA} N_{caseDB}} \,.$$

Simplifying this,

$$\sigma_{uDA,uDB} = \frac{M_{TA} N_{caseDA}}{(1 - M_{TA}) N_{caseDB}} \,,$$

and expressing in terms of M_{TA} yields the relationship

$$\frac{\sigma_{\mathrm{uDA,uDB}} N_{\mathrm{caseDB}}}{N_{\mathrm{caseDA}}} = \frac{M_{\mathrm{TA}}}{(1 - M_{\mathrm{TA}})} \quad \text{(Equation 10)}$$

Substituting $C = \sigma_{\mathrm{uDA,uDB}} N_{\mathrm{caseDB}} / N_{\mathrm{caseDA}}$ (the left-hand term) and solving for M_{TA} gives the relationship

$$M_{\mathrm{TA}} = \frac{C}{C+1} \quad \text{(Equation 11)}$$

Using mixed linear model GWAS association statistics for ALS and constrained intercept LD score regression, the genetic correlation (r_g) estimate for ALS and schizophrenia was 14.3% (7.05-21.6). Using the relationship

$$\rho_{\mathrm{g,l}} = r_{\mathrm{g}} \sqrt{h^2_{\mathrm{ALS,l}} h^2_{\mathrm{SCZ,l}}} \quad \text{(ref 11)}$$

where subscript *l* denotes liability-scale, this corresponds to a liability-scale genetic covariance of 1.88% (0.92-2.83). With GWAS case cohort sizes of N_{caseDA} = 12,577 and N_{caseDB} = 34,241, M_{TA} evaluates to 4.86% (2.47-7.13). Supplementary table 7 shows M_{TA} estimates for the study's other $\rho_{g,l}$ estimates.

6.5.2 *Supplementary methods 2: Modelling expected comorbidity*

Under a positive genetic correlation between two binary phenotypes, an increased number of individuals exhibiting both phenotypes would be expected. To determine the required cohort size to observe a significant excess of comorbidity, we modelled the bivariate distribution in liability for pairs of traits as $N_2(0, \Sigma)$ given the covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho_{\mathrm{g,l}} \\ \rho_{\mathrm{g,l}} & 1 \end{bmatrix},$$

where

$$\rho_{\mathrm{g,l}} = r_{\mathrm{g}} \sqrt{h^2_{\mathrm{ALS,l}} h^2_{\mathrm{SCZ,l}}}$$

is the liability-scale genetic covariance. We calculated the liability thresholds t_A and t_B for traits A and B as the Normal quantiles above which proportions P_A and P_B of liability lie, corresponding to lifetime risks for traits A and B, respectively. We then calculated the expected proportion P_{AB} of cases that pass these thresholds from the bivariate cumulative distribution of liability using the algorithm of Genz *et al.* [41] implemented in the *mvtnorm* package in R 3.0.2.

The odds ratio for developing ALS given a diagnosis of schizophrenia is (Supplementary Figure 4):

$$OR_{\mathrm{ALS|SCZ}} = \frac{P_{\mathrm{AB}}/(P_{\mathrm{B}} - P_{\mathrm{AB}})}{(P_{\mathrm{A}} - P_{\mathrm{AB}})/(1 - P_{\mathrm{A}} - P_{\mathrm{B}} - P_{\mathrm{AB}})} . \quad \text{(Equation 12)}$$

The odds ratio for developing schizophrenia given a diagnosis of ALS is

$$OR_{\mathrm{SCZ|ALS}} = \frac{P_{\mathrm{AB}}/(P_{\mathrm{A}} - P_{\mathrm{AB}})}{(P_{\mathrm{B}} - P_{\mathrm{AB}})/(1 - P_{\mathrm{A}} - P_{\mathrm{B}} - P_{\mathrm{AB}})} . \quad \text{(Equation 13)}$$

Cross-multiplying, both equations become

$$OR = \frac{P_{\mathrm{AB}}(1 - P_{\mathrm{A}} - P_{\mathrm{B}} - P_{\mathrm{AB}})}{(P_{\mathrm{A}} - P_{\mathrm{AB}})(P_{\mathrm{B}} - P_{\mathrm{AB}})} . \quad \text{(Equation 14)}$$

Under no genetic correlation the expected proportion P_0 of individuals passing liability thresholds t_A and t_B for traits A and B is the bivariate cumulative distribution function for $N_2(0,\mathbf{I})$ at t_A and t_B, which is equivalent to $P_A P_B$ (**I** is the identity matrix). This was used to calculate the size of population *n* required to achieve 80% power ($\beta = 0.2$) to observe a significant ($\alpha = 0.05$) excess of co-occurrences of the two traits using the two-tailed binomial power equation,

$$\mathrm{n} = P_{\mathrm{AB}}(1 - P_{\mathrm{AB}})\left(\frac{\phi^{-1}_{1-\frac{\alpha}{2}} + \phi^{-1}_{1-\beta}}{P_{\mathrm{AB}} - P_0}\right)^2 , \quad \text{(Equation 15)}$$

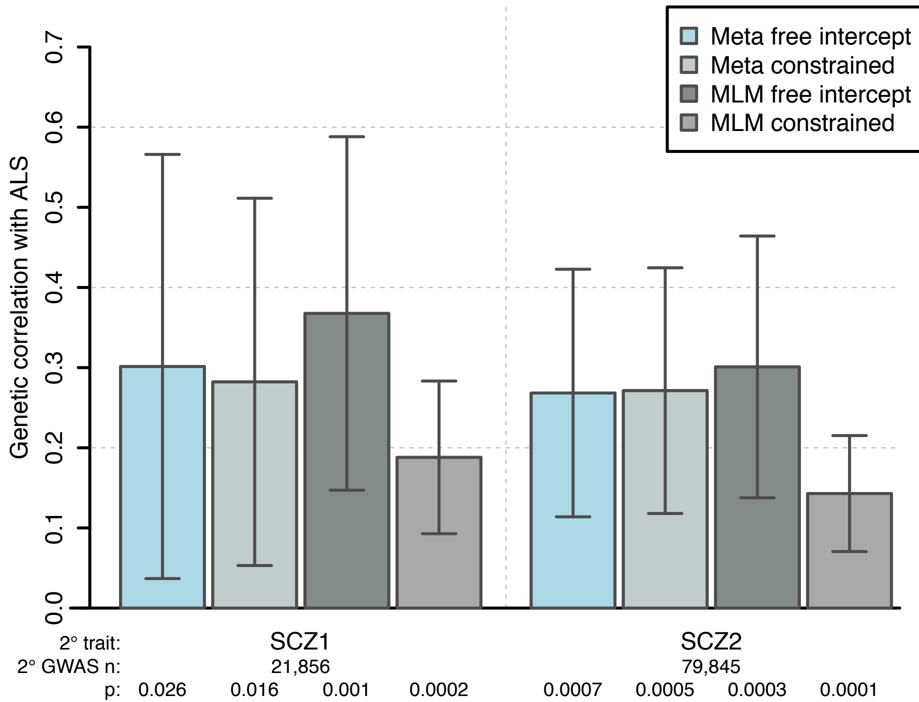
Where Φ is the cumulative distribution function of the Normal distribution. Multiplying *n* by the lifetime risk for developing ALS (1/400) gives the required size of incident ALS cohort to observe this excess.

6.5.3 Supplementary note 1: IGAP data

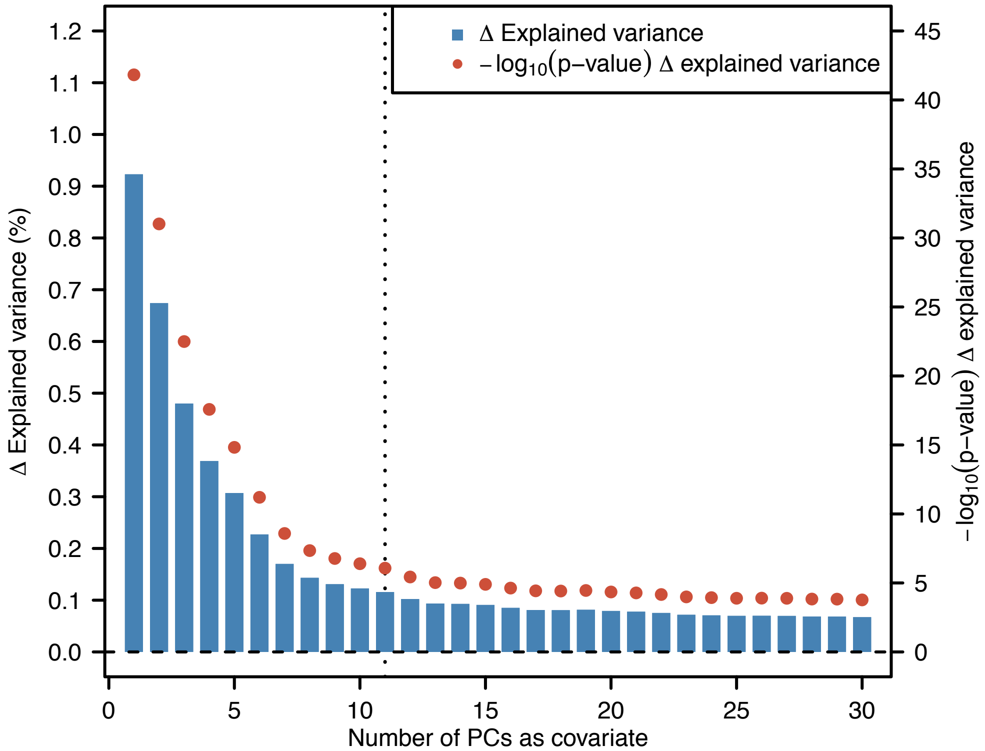
This study made use of data from the International Genomics of Alzheimer's Project (IGAP), a large two-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 7,055,881 single nucleotide polymorphisms (SNPs) to meta-analyse four previously- published GWAS datasets consisting of 17,008 Alzheimer's disease cases and 37,154 controls (The European Alzheimer's disease Initiative – EADI the Alzheimer Disease Genetics Consortium – ADGC The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium – CHARGE The Genetic and Environmental Risk in AD consortium – GERAD). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and

11,312 controls. Finally, a meta-analysis was performed combining results from stages 1 & 2.

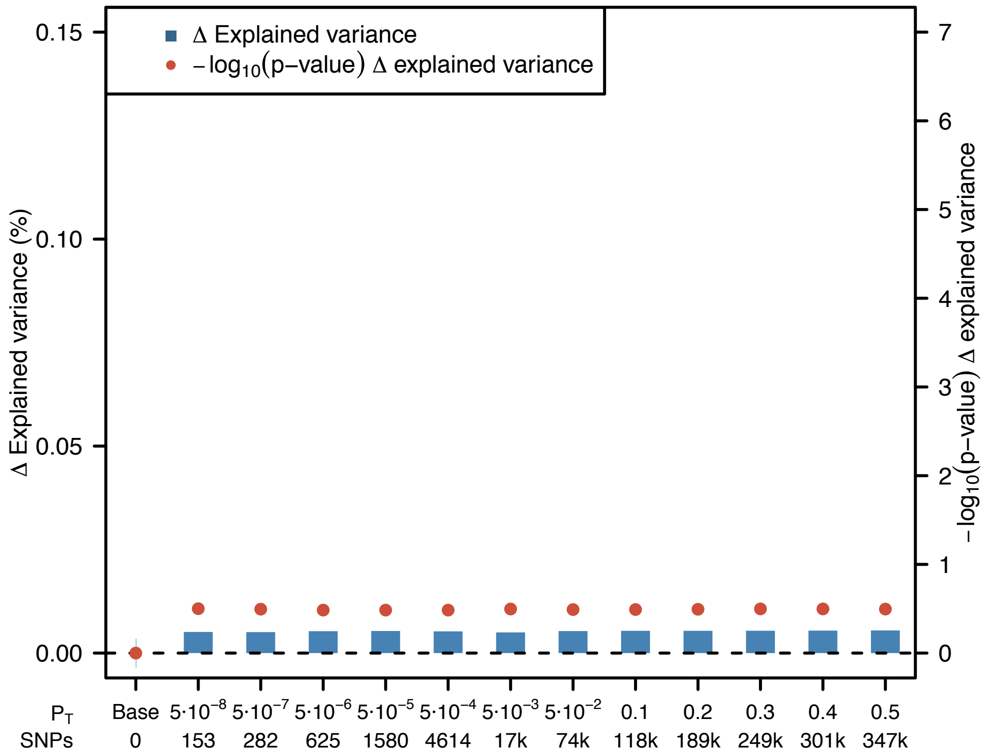
6.5.4 Supplementary figures



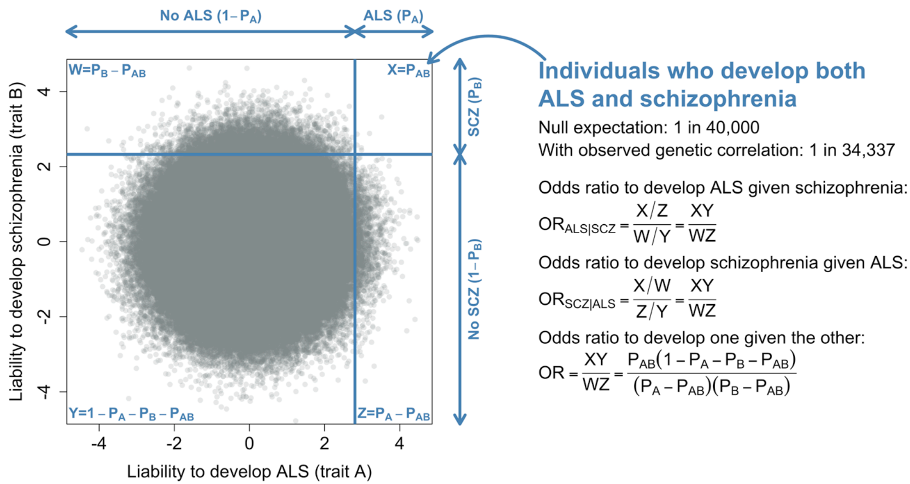
Supplementary Figure 1. Comparison of genetic correlation estimates for two schizophrenia cohorts. SCZ1 denotes data from the 2011 study by the Schizophrenia Psychiatric Genome-Wide Association Study Consortium (European ancestry cohort); SCZ2 indicates data from the 2014 study by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (European and Asian ancestry cohort). Error bars indicate 95% confidence intervals.



Supplementary Figure 2. Effect of increasing the number of PCs as covariates in generalized linear model on explained variance in ALS and -$\log_{10}$(p-values) attributable to polygenic risk scores (Δ explained variance) at schizophrenia P_T = 0.2. The dotted line indicates the 11 most significant PCs ($p < 0.0005$) used in the main analysis.



Supplementary Figure 3. Effect of 1000 permutations of case-control labels, with conservation of case:control ratio, on analysis of polygenic risk scores (PRS) for schizophrenia in the ALS dataset. The schizophrenia P_T and associated number of SNPs are similar to the main analysis. Blue bars show the averaged Δ explained variance of a linear model containing PRS compared to a baseline model with only covariates. Red dots indicate averaged $-\log_{10}$(p-values) of Δ explained variance.



$$OR_{ALS|SCZ} = \frac{X/Z}{W/Y} = \frac{XY}{WZ}$$

$$OR_{SCZ|ALS} = \frac{X/W}{Z/Y} = \frac{XY}{WZ}$$

$$OR = \frac{XY}{WZ} = \frac{P_{AB}(1 - P_A - P_B - P_{AB})}{(P_A - P_{AB})(P_B - P_{AB})}$$

Supplementary Figure 4. Simulated example of the bivariate distribution of liability for two traits with a genetic covariance (liability-scale) of 1.88% in a population of 500,000 individuals (grey dots). P_A and P_B represent the proportions of the population that carry above-threshold liabilities for traits A and B (defined by the lifetime risks for these traits). Individuals in sector X in the plot carry above-threshold liability for both traits A and B (P_{AB} of the population).

6.5.5 Supplementary tables

Supplementary Table 1. LD score regression heritability estimates for amyotrophic lateral sclerosis.

Method	h_s^2(%)	95% CI
MLM constrained intercept	8.17	7.21-9.13
MLM free intercept	1.54	0.13-2.95
Meta constrained intercept	1.34	0.61-2.07
Meta free intercept	1.61	0.67-2.55
Permuted constrained intercept	-0.7	-1.50-0.18
Permuted free intercept	0.09	-1.09-1.27

h_s^2, SNP-based heritability; CI, confidence interval; MLM, mixed linear model summary statistics; meta, meta-analysis summary statistics

Supplementary Table 2. Genetic correlation between amyotrophic lateral sclerosis and 8 secondary traits.

	2° Trait	AD	ADHD	ASD	BPD	HEIGHT	MDD	MS	SCZ
	GWAS n	54,162	5,422	10,263	16,731	253,288	18,759	17,597	79,845
	K	0.15	0.05	0.01	0.01	n/a	0.15	0.0029	0.01
	h_S^2 % (s.e.)	8.1 (1.6)	9.5 (6.7)	17.0 (2.3)	17.1 (2.1)	27.4 (1.6)	13.4 (3.1)	19.1 (7.8)	19.7 (1.0)
MLM cons	ρ_g % (s.e.)	0.5 (0.8)	-0.7 (1.9)	-1.0 (1.0)	-0.2 (0.8)	-0.7 (0.4)	0.02 (1.0)	0.3 (0.6)	1.8 (0.5)
	r_g % (s.e.)	6.3 (9.6)	-8.3 (21.9)	-8.6 (8.7)	-1.9 (6.6)	-4.5 (2.9)	0.2 (9.8)	2.1 (4.7)	14.3 (3.7)
	p	0.51	0.71	0.32	0.78	0.11	0.98	0.65	**0.0001**
MLM free	ρ_g % (s.e.)	1.5 (1.8)	-1.6 (4.0)	-3.1 (2.2)	-0.3 (1.6)	-1.4 (1.0)	-0.3 (1.9)	0.9 (1.4)	3.8 (1.1)
	r_g % (s.e.)	18.8 (22.4)	-18.4 (44.7)	-25.8 (19.0)	-2.9 (13.4)	-9.2 (6.7)	-2.8 (18.3)	6.9 (11.0)	30.1 (8.3)
	p	0.40	0.68	0.18	0.83	0.18	0.88	0.53	**0.0003**
Meta cons	ρ_g % (s.e.)	1.9 (1.8)	-0.9 (3.2)	-2.3 (2.1)	-0.9 (1.5)	-0.2 (0.9)	0.1 (2.0)	1.0 (1.2)	3.4 (1.0)
	r_g % (s.e.)	23.3 (21.8)	-10.0 (36.3)	-19.1 (18.0)	-8.0 (12.6)	-1.3 (6.0)	1.3 (18.9)	8.3 (9.8)	27.1 (7.8)
	p	0.28	0.78	0.29	0.52	0.83	0.94	0.40	**0.0005**
Meta free	ρ_g % (s.e.)	2.0 (1.8)	-0.9 (3.6)	-2.4 (2.2)	-1.1 (1.7)	-0.2 (0.9)	0.1 (2.2)	1.2 (1.3)	3.4 (1.0)
	r_g % (s.e.)	24.5 (21.6)	-10.4 (40.5)	-20.1 (19.0)	-9.1 (14.2)	-1.4 (6.3)	0.9 (21.1)	9.2 (10.5)	26.8 (7.9)
	p	0.26	0.80	0.27	0.52	0.83	0.97	0.38	**0.0007**

AD, Alzheimer's disease; ADHD, attention deficit-hyperactivity disorder; ASD, autism spectrum disorder; BPD, bipolar disorder; MDD, major depressive disorder; MS, multiple sclerosis; SCZ, schizophrenia; n, sample size; K, population prevalence (lifetime risk) for binary traits; h_S^2, SNP-based heritability; ρ_g, genetic covariance; r_g, genetic correlation; s.e., standard error; MLM, mixed linear model summary statistics for ALS; Meta, meta-analysis summary statistics for ALS; cons, constrained-intercept LD score regression for primary trait (ALS) h_S^2 estimation

p-values are emboldened if they are significant at (conservative) Bonferroni-corrected $\alpha = 0.0016$

h_S^2 and ρ_g estimates are liability-scale for binary traits

h_S^2 estimates for secondary traits used free-intercept LD score regression

Supplementary Table 3. Number of individuals in the ALS dataset excluded from PRS analysis due to (suspected) sample overlap with the schizophrenia dataset.

Stratum	Name	Total individuals	Excluded individuals	Percentage
1	NL1	843	5	0.6
2	BE1	616	1	0.2
3	NL2	5027	25	0.5
4	SW1	556	82	14.7
5	FR1	809	0	0.0
6	UK1	327	0	0.0
7	US1	1937	4	0.2
8	IR1	639	16	2.5
9	UK2	3301	2555	77.4
10	US2	779	4	0.5
11	IT1	626	0	0.0
12	FIN1	756	756	100.0
13	NL3	2781	86	3.1
14	GER1	776	776	100.0
15	IT2	383	0	0.0
16	IB1	225	0	0.0
17	SWISS1	424	0	0.0
18	BE2	447	1	0.2
19	SW2	467	60	12.8
20	FIN2	232	232	100.0
21	IR2	707	299	42.3
22	US3	2562	3	0.1
23	FR2	1332	0	0.0
24	UK3	3534	2433	68.8
25	GER2	2047	2047	100.0
26	IT3	2790	0	0.0
27	NL4	1129	8	0.7
	Total	36052	9393	26.1

NL, Netherlands; BE, Belgium; SW, Sweden; FR, France; UK, United Kingdom; US; United States of America; IR, Ireland; IT, Italy; FIN, Finland; GER, Germany; IB, Iberia (Spain and Portugal); SWISS, Switzerland.

Supplementary Table 4. Genomic regions with high LD removed prior to polygenic risk score calculation.

Chromosome	Basepair start	Basepair end
1	48000000	52000000
2	86000000	100500000
	183000000	190000000
3	47500000	50000000
	83500000	87000000
5	44500000	50500000
	129000000	132000000
6	25500000	33500000
	57000000	64000000
	140000000	142500000
7	55000000	66000000
8	8000000	12000000
	43000000	50000000
	8135000	12000000
	112000000	115000000
10	37000000	43000000
11	87500000	90500000
12	33000000	40000000
17	40900000	45000000
20	32000000	34500000

Positions refer to human genome build GRCh37

Supplementary Table 5. Estimated explained variances and p-values of polygenic risk scores through generalized linear model.

SCZ P_T	n SNPs	Explained variance (Nagelkerke R^2, %)	Δ Explained variance (Nagelkerke R^2, %)	p-value	$-log_{10}$(p-value)
Baseline	0	9.050	-	-	-
5×10^{-8}	153	9.051	0.001	6.90×10^{-1}	0.16
5×10^{-7}	282	9.056	0.006	2.73×10^{-1}	0.56
5×10^{-6}	625	9.053	0.002	5.10×10^{-1}	0.29
5×10^{-5}	1,580	9.057	0.007	2.36×10^{-1}	0.63
5×10^{-4}	4,614	9.081	0.030	1.15×10^{-2}	1.94
5×10^{-3}	17,042	9.126	0.075	7.15×10^{-5}	4.15
0.05	74,036	9.143	0.093	1.01×10^{-5}	4.99
0.1	118,335	9.163	0.112	1.27×10^{-6}	5.90
0.2	189,492	9.166	0.116	8.43×10^{-7}	6.07
0.3	249,452	9.162	0.111	1.41×10^{-6}	5.85
0.4	301,405	9.164	0.114	1.06×10^{-6}	5.97
0.5	346,683	9.161	0.111	1.50×10^{-6}	5.82

SCZ, schizophrenia; P_T, p-value threshold

Supplementary Table 6. ALS case-control ratio for schizophrenia polygenic risk score deciles.

Decile	Cases	Controls	OR vs decile 1	95% CI	Z value	p-value
1	732	1,934	-	-	-	-
2	801	1,865	1.02	0.90-1.15	0.27	7.6×10^{-1}
3	882	1,784	1.11	0.98-1.25	1.60	9.4×10^{-2}
4	910	1,756	1.10	0.97-1.24	1.47	1.3×10^{-1}
5	967	1,698	1.13	1.00-1.27	1.89	4.8×10^{-2}
6	982	1,685	1.10	0.97-1.24	1.39	1.3×10^{-1}
7	1,056	1,610	1.16	1.03-1.31	2.26	1.7×10^{-2}
8	1,103	1,562	1.15	1.02-1.30	2.06	2.6×10^{-2}
9	1,224	1,442	1.25	1.10-1.41	3.35	3.9×10^{-4}
10	1,375	1,291	1.30	1.15-1.48	3.84	5.1×10^{-5}

OR, Odds ratio, CI, confidence interval

Supplementary Table 7. Misdiagnosis rates (M_{TA}) of ALS as schizophrenia required to yield genetic correlation estimates assuming true genetic correlation of 0.

Method	r_g (%)	95% CI	$\rho_{g,l}$ (%)	95% CI	M_{TA} (%)	95% CI
MLM constrained	14.3	7.08-21.6	1.88	0.92-2.83	4.86	2.47-7.13
MLM free	30.1	13.8-46.4	3.95	1.81-6.09	9.70	4.70-14.2
Meta constrained	27.1	11.8-42.4	3.55	1.55-5.56	8.82	4.04-13.1
Meta free	26.8	11.4-42.2	3.52	1.50-5.53	8.73	3.91-13.1

r_g, genetic correlation between ALS and schizophrenia; CI, confidence interval; $\rho_{g,l}$, liability-scale genetic covariance; M_{TA} misdiagnosis rate

6

Supplementary Table 8. Genomic loci associated with ALS at $cFDR_{ALS|SCZ}$ < 0.05.

Index SNP	Chromosome	Position (GRCh37)	GWAS p_{ALS}	GWAS p_{SCZ}	$cFDR_{ALS|SCZ}$	Closest gene
rs63535762	1	233663807	4.0882×10^{-6}	0.1232	0.02914227	*KCNK1*
rs11124852	2	42131864	5.6779×10^{-6}	0.04953	0.03113055	*C2orf91*
rs149853584	3	1728150	3.5418×10^{-8}	0.03095	0.00038322	*CNTN6*
rs2625891	3	9541761	2.763×10^{-6}	0.03099	0.01459386	*LHFPL4*
rs616147	3	39534481	5.7276×10^{-8}	0.6211	0.00133118	*MOBP*
rs1817362	3	147770849	8.3621×10^{-6}	0.009266	0.03495604	*AGTR1*
rs71308531	3	171681860	2.0263×10^{-6}	0.5228	0.02195072	*FDNC3B*
rs13164762	5	83141157	4.5304×10^{-6}	0.07264	0.02709394	*EDIL3*
rs34384833	5	91238155	9.9144×10^{-6}	0.0399	0.04847741	*ARRDC3*
rs189841212	5	117878641	6.5395×10^{-6}	0.02757	0.02927102	*DTWD2*
rs10463311	5	150410835	3.645×10^{-7}	0.1378	0.00317718	*TNIP1*
rs116946806	7	131682571	2.5608×10^{-6}	0.3095	0.02134998	*PLXNA4*
rs17070492	8	2420855	4.0231×10^{-6}	0.01049	0.01994008	*MYOM2*
rs7814166	8	143370923	6.8943×10^{-5}	1.373×10^{-10}	0.0427745	*TSNARE1*
rs10965525	9	22936164	5.0503×10^{-6}	0.006832	0.02977711	*DMRTA1*
rs2484319	9	27563755	2.0583×10^{-17}	0.007078	2.8215×10^{-12}	*C9orf72*
rs76794160	9	71621214	1.5665×10^{-6}	0.1176	0.01059827	*PIP5K1B*
rs11146348	10	133772553	7.1712×10^{-7}	0.0113	0.00453269	*PPP2R2D*
rs143276647	11	118299359	8.2063×10^{-6}	0.06059	0.03712072	*KMT2A*
rs79676202	12	50180558	4.5475×10^{-8}	0.369	0.00070155	*NCKAP5L*
rs113247976	12	57975700	4.1051×10^{-6}	0.1964	0.03831843	*KIF5A*
rs116900480	12	58656105	1.4775×10^{-6}	0.1303	0.01195632	*XRCC6BP1*
rs76805704	12	64532377	1.5297×10^{-6}	0.1331	0.01220552	*SRGAP1*
rs74654358	12	64881967	1.5956×10^{-7}	0.02022	0.00103714	*TBK1*
rs179552	14	31225831	9.8359×10^{-7}	0.1131	0.01156173	*SCFD1*
rs3097439	15	27871229	7.5601×10^{-6}	0.1372	0.0401236	*GABRG3*
rs35714695	17	26719788	3.1748×10^{-9}	0.7655	0.00023563	*SARM1*
rs78549703	19	17749542	1.6827×10^{-9}	0.1015	1.4414×10^{-5}	*UNC13A*
rs6015322	20	57206540	1.3812×10^{-6}	0.4992	0.03610187	*APCDD1L-AS1*
rs117635456	21	43460912	4.377×10^{-7}	0.3587	0.00546115	*ZNF295-AS1*
rs2176039	22	45585032	9.1138×10^{-6}	0.02482	0.04064275	*NUP50*

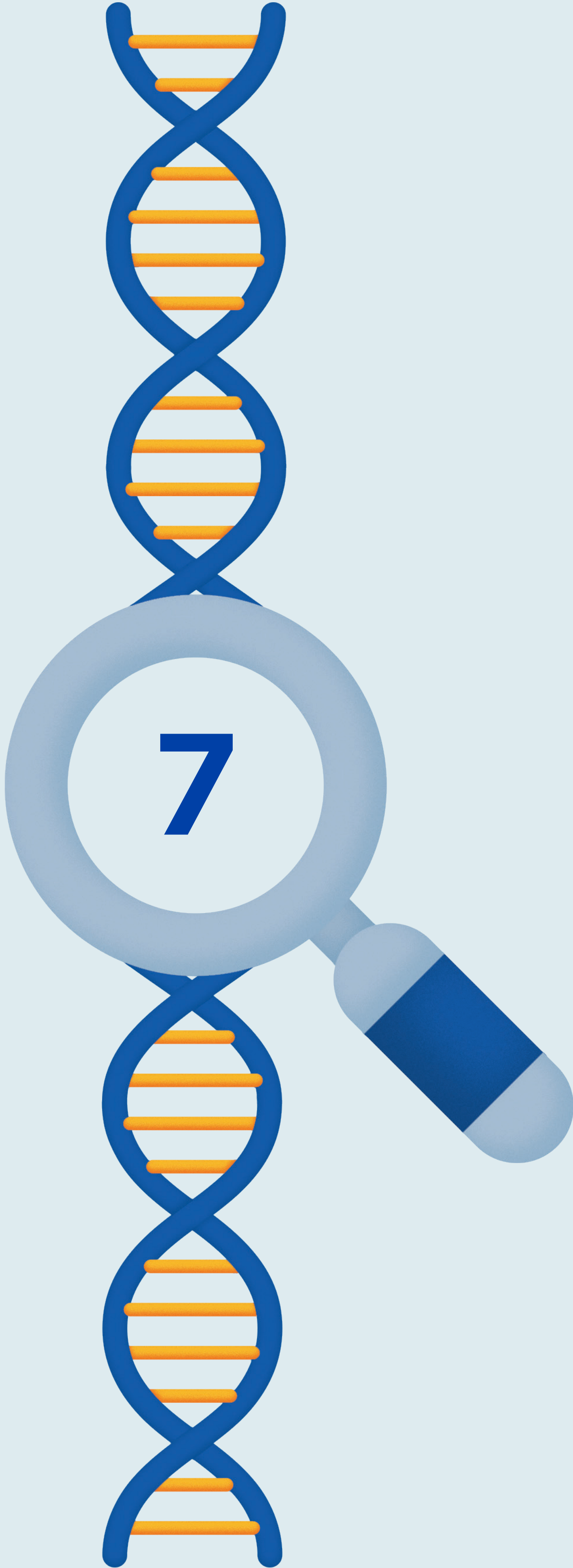
SNP, single nucleotide polymorphism; GWAS, genome-wide association study; ALS, amyotrophic lateral sclerosis; SCZ, schizophrenia; cFDR, conditional false discovery rate

6.6 References

1. Phukan J, Elamin M, Bede P, Jordan N, Gallagher L, Byrne S, et al. The syndrome of cognitive impairment in amyotrophic lateral sclerosis: a population-based study. J Neurol Neurosurg Psychiatry 2012;83(1):102–8.
2. Byrne S, Heverin M, Elamin M, Bede P, Lynch C, Kenna K, et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: A population-based case–control cohort study of familial and sporadic amyotrophic lateral sclerosis. Ann Neurol 2013;74(5):699–708.
3. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J Neurol Neurosurg Psychiatry 2010;81(12):1324–6.
4. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 2009;373(9659):234–9.
5. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511(7510):421–7.
6. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013;45(9):984–94.
7. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet 2016;48(9):1043–8.
8. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.
9. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, et al. Improved Detection of Common Variants Associated with Schizophrenia and Bipolar Disorder Using Pleiotropy-Informed Conditional False Discovery Rate. PLoS Genet 2013;9(4):e1003455.
10. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460(7256):748–52.
11. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet 2014;46(2):100–6.
12. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190–1.
13. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Psychiatric GWAS Consortium Bipolar Disorder Working Group, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47(3):291–5.
14. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011;43(10):969–76.
15. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 2011;43(10):977–83.
16. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013;18(4):497–511.
17. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013;381(9875):1371–9.
18. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45(12):1452–8.

19. Patsopoulos NA, the Bayer Pharma MS Genetics Working Group, the Steering Committees of Studies Evaluating IFNβ-1b and a CCR1-Antagonist, ANZgene Consortium, GeneMSA, International Multiple Sclerosis Genetics Consortium, de Bakker PIW. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Ann Neurol 2011;70(6):897–912.
20. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46(11):1173–86.
21. Han B, Pouget JG, Slowikowski K, Stahl E, Lee CH, Diogo D, et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nat Genet 2016;48(7):803–10.
22. Liley J, Wallace C. A Pleiotropy-Informed Bayesian False Discovery Rate Adapted to a Shared Control Design Finds New Disease Associations From GWAS Summary Statistics. PLoS Genet 2015;11(2):e1004926.
23. Brown AS, McGrath JJ. The prevention of schizophrenia. Schizophr Bull 2011;37(2):257–61.
24. Al-Chalabi A, Hardiman O, Kiernan MC, Chiò A, Rix-Brooks B, van den Berg LH. Amyotrophic lateral sclerosis: moving towards a new classification system. Lancet Neurol 2016;15(11):1182–94.
25. Kahn RS, Sommer IE, Murray RM, Meyer-Lindenberg A, Weinberger DR, Cannon TD, et al. Schizophrenia. Nat Rev Dis Primers 2015;:15067.
26. Huang Z, Yu Y, Shimoda Y, Watanabe K, Liu Y. Loss of neural recognition molecule NB-3 delays the normal projection and terminal branching of developing corticospinal tract axons in the mouse. J Comp Neurol 2012;520(6):1227–45.
27. Ramirez VP, Gurevich I, Aneskievich BJ. Emerging roles for TNIP1 in regulating post-receptor signaling. Cytokine Growth Factor Rev 2012;23(3):109–18.
28. Toker A, Marmiroli S. Signaling specificity in the Akt pathway in biology and disease. Adv Biol Regul 2014;55:28–38.
29. Wang K-S, Liu X-F, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophr Res 2010;124(1-3):192–9.
30. Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. PLoS Genet 2012;8(4):e1002635.
31. Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 2004;36(1):40–5.
32. Laursen TM, Munk-Olsen T, Vestergaard M. Life expectancy and cardiovascular mortality in persons with schizophrenia. Curr Opin Psychiatry 2012;25(2):83–8.
33. Stommel EW, Graber D, Montanye J, Cohen JA, Harris BT. Does treating schizophrenia reduce the chances of developing amyotrophic lateral sclerosis? Med Hypotheses 2007;69(5):1021–8.
34. Farokhnia M, Sabzabadi M, Pourmahmoud H, Khodaie-Ardakani M-R, Hosseini S-M-R, Yekehtaz H, et al. A double-blind, placebo controlled, randomized trial of riluzole as an adjunct to risperidone for treatment of negative symptoms in patients with chronic schizophrenia. Psychopharmacology (Berl) 2014;231(3):533–42.
35. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56–65.
36. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2(12):e190.
37. Wray NR, Lee SH, Kendler KS. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. Eur J Hum Genet 2012;20(6):668–74.
38. Johnston CA, Stanton BR, Turner MR, Gray R, Blunt AH-M, Butt D, et al. Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. J Neurol 2006;253(12):1642–3.

39. Andreassen OA, Harbo HF, Wang Y, Thompson WK, Schork AJ, Mattingsdal M, et al. Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. Mol Psychiatry 2015;20(2):207–14.
40. Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D, DIAGRAM Consortium, et al. Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. PLoS Genet 2010;6(8):e1001058.
41. Genz A. Numerical Computation of Multivariate Normal Probabilities. J Comput Graph Stat 1992;1(2):141.


7

Chapter 7

Common genetic risk is not shared between amyotrophic lateral sclerosis and epilepsy

Dick Schijven, Remi Stevelink, Mark McCormack, Wouter van Rheenen, Jurjen J. Luykx, Bobby P.C. Koeleman, Jan H. Veldink, on behalf of the Project MinE ALS GWAS Consortium[#] and the International League Against Epilepsy Consortium on Complex Epilepsies[#]

[#]Consortium members are listed in appendix I

Manuscript in preparation

Abstract

Since hyperexcitability has been shown to be a shared pathophysiological mechanism between amyotrophic lateral sclerosis and epilepsy, we set out to determine genetic overlap between these conditions. We used the latest and largest genome-wide studies on amyotrophic lateral sclerosis (N = 36,052) and epilepsy (N = 38,349). First, we showed no significant genetic correlation, also when binned on minor allele frequency. In addition, the genetic architectures of ALS and epilepsy were found to differ. Second, we confirmed the absence of polygenic overlap using genomic risk score analysis. Finally, no credible pleiotropic variants were found in meta-analyses of the two diseases.

7.1 Introduction

Amyotrophic lateral sclerosis (ALS) is characterized by the progressive degeneration of motor neurons. Cramps and fasciculations are among the most observed symptoms in ALS patients and are the result of axonal hyperexcitability in motor neurons caused by increased sodium currents followed by decreased potassium currents [1]. Peripheral and central hyperexcitability are furthermore strong predictors of shorter survival in ALS [2,3]. Epilepsy is a common neurological disorder in which an imbalance of excitatory and inhibitory mechanisms in the brain results in a predisposition to develop seizures. Interestingly, the anti-epileptic drug retigabine has shown acute beneficial effects on peripheral nerve excitability in ALS [4].

These clinical observations and the availability of recent and large genome-wide association study (GWAS) datasets [5,6], which showed that the heritability of both diseases is partly explained by common genetic variants, have led us to investigate whether ALS and epilepsy share a certain degree of common genetic risk.

7.2 Materials and methods

7.2.1 Study population and data

We used individual level data of subjects with European ancestry from recent large GWASs on ALS (N = 12,577 cases; 23,475 controls) [5] and epilepsy (N = 14,131 cases; 24,218 controls) [6], including subsets of focal (n = 9,095 cases) and generalized (n = 3,305 cases) subtypes. Written informed consent was given by all participants and for both studies local institutional review boards at each contributing site approved study protocols [5,6]. In all analyses single-nucleotide polymorphisms (SNPs) with a genotype call rate < 0.95 were excluded, except when stated otherwise.

7.2.2 Genetic correlation and heritability analyses

We merged individual level data of ALS and epilepsy using PLINK (version v1.90b5.4) [7], excluded SNPs with a genotype call rate < 0.99, SNPs not present in HapMap3 and SNPs located in the major histocompatibility complex (MHC) region. Five genetic relationship matrices (GRMs) binned on minor allele

7

frequencies (MAFs) based on the European subset of 1000 genomes phase 3 (1KG-EUR, n = 503) (0.01-0.10, 0.10-0.20, 0.20-0.30, 0.30-0.40, 0.40-0.50) were calculated using Genome-wide Complex Trait Analysis (GCTA, version 1.91) [8]. We used bivariate genetic restricted maximum likelihood (GREML) to calculate genetic correlation (r_g) per MAF bin and on all bins combined. In support of this, we estimated overall r_g from summary-level data using linkage disequilibrium score regression (LDSC) with default settings [9]. In addition, we calculated MAF binned GRMs for ALS and epilepsy separate (excluding SNPs with a genotyping rate < 0.99 and SNPs located in the MHC region) and estimated the proportion of SNP-based heritability on the liability scale (h_S^2) explained by each bin [10] using univariate GREML and disease prevalences of 0.0025 (ALS), 0.006 (all epilepsy), 0.003 (focal epilepsy) and 0.002 (generalized epilepsy) [11-13].

7.2.3 Genomic risk score analysis

Genomic risk scores (GRSs) reflect the sum of discovery GWAS SNP effect alleles, weighted for their effect size, in individuals in a target dataset. GRSs therefore capture the combined effect of many SNPs and can be tested for association with a phenotype in the target data. We used epilepsy mixed linear model (MLM) GWAS results as discovery data and ALS as a target dataset, and vice versa, excluding individuals from the target dataset that were related to individuals used to generate the discovery data (PLINK identity-by-descent PIHAT > 0.10). In the discovery data, we excluded variants other than SNPs, and SNPs with MAF < 0.01, a strand-ambiguous A/T or C/G genotype and those located in genomic regions with a strong or complex LD structure (Supplementary Table 1). LD-independent variants were obtained through clumping of SNPs that intersected with the target dataset using 1KG-EUR as a reference for LD estimation and subsequent pruning of SNPs with an LD $R^2 > 0.2$ with the strongest associated variant in a 5000 kilobase radius. Genomic risk scores (GRSs) were calculated using PLINK and scaled around mean 0 with variance 1. We applied binomial regression to test association between disease phenotype and GRSs.

7.2.4 Meta-analysis

We replicated a previously reported MLM association analysis for ALS using GCTA [5], now excluding duplicate and related individuals with the epilepsy

data (PLINK PIHAT > 0.10). Inverse-variance weighted fixed effects meta-analyses were performed in METAL [14] for MLM GWAS results of ALS and epilepsy (including subtypes). Similarly, we performed meta-analyses where we ignored SNP effect direction to discover pleiotropic SNPs with opposing effect directions.

7.3 Results

We combined ALS and epilepsy GWAS datasets (totaling 74,401 individuals), applied genetic correlation, heritability and GRS analyses to assess shared heritability between these diseases, and performed meta-analyses to discover possible pleiotropic loci.

No significant genetic correlations were found using GCTA-GREML between ALS and epilepsy (r_g = 0.0080 [95 % confidence interval -0.079–0.095], p = 0.86), including focal subtype (r_g = 0.052 [-0.046–0.15], p = 0.29) and generalized subtype (r_g = -0.011 [-0.099–0.077], p = 0.81), neither when the analysis was stratified on MAF (Figure 1A). The absence of r_g with ALS was confirmed using LDSC for all (r_g = 0.054 [-0.23–0.34], p =0.70), focal (r_g = 0.49 [-0.060–1.04], p = 0.08), and generalized epilepsy (r_g = -0.055 [-0.31–0.20], p = 0.67) (Figure 1A). With the currently available datasets, binned heritability analysis showed the trend that in ALS a larger proportion (0.52) of the heritability was explained by SNPs with $0.01 \leq MAF < 0.10$ compared to other bins, while the heritability of epilepsy was equally distributed across bins (Figure 1B). Furthermore, GRSs capturing genetic risk for epilepsy did not significantly add to the phenotypic variance explained in ALS (Figure 2A, Supplementary Table 2A), and similarly ALS GRSs did not explain phenotypic variance in epilepsy (Figure 2B, Supplementary Table 2B).

Inverse-variance weighted meta-analyses of ALS and epilepsy revealed several loci passing the threshold for genome-wide significance (Figure 3, Supplementary Figure 1, Supplementary Table 3A). Only the locus on chromosome 1p34.3 near *BMP8A* in the ALS-epilepsy (all subtypes combined) analysis (Figure 3A) fulfilled criteria for a pleiotropic locus with nominal-significant association p-values ($p < 0.05$) in both studies independently (ALS $p = 2.3 \times 10^{-2}$, epilepsy $p = 8.7 \times 10^{-8}$) and a genome-wide significant

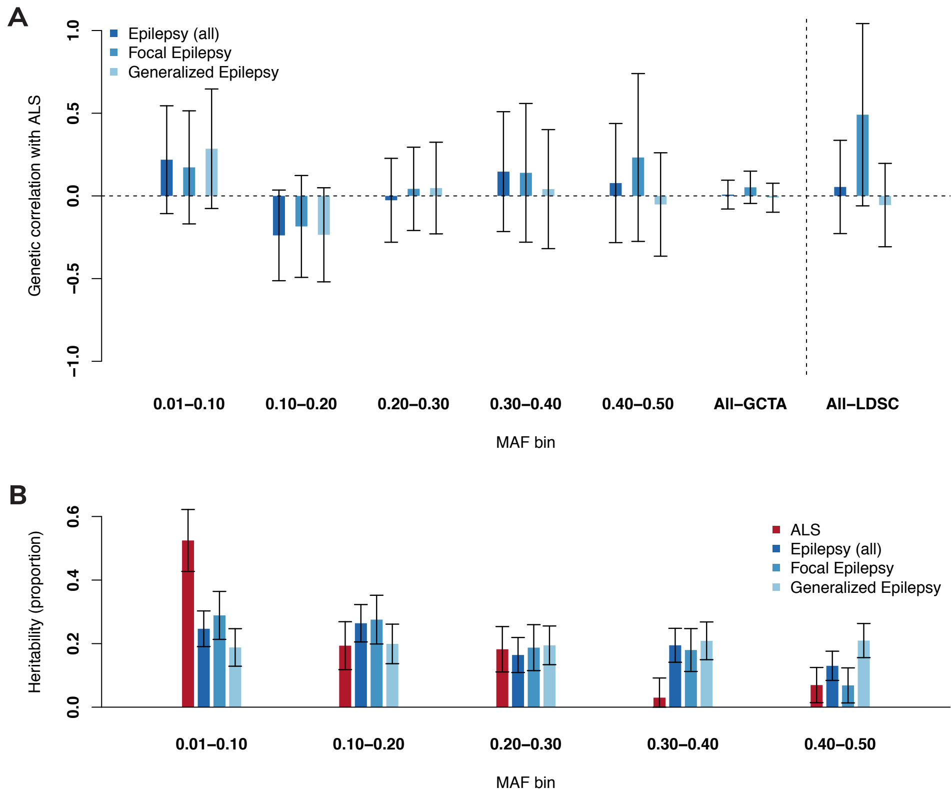


Figure 1. Genetic correlation and SNP-based heritability between ALS and epilepsy. A) Genetic correlation between ALS and epilepsy (including focal and generalized subtypes) (y-axis) was estimated in five MAF bins and all bins combined using GCTA-REML (x-axis). Estimates were calculated in unrelated individuals (genetic relationship matrix (GRM) off-diagonal < 0.05) and the first 10 genetic principal components (PCs) derived from GRMs were used as covariates. Furthermore, LDSC was used to estimate genetic correlation (All-LDSC). LD scores were calculated in 1 centiMorgan windows around SNPs present in HapMap3, using the European subset of 1000 genomes (1KG-EUR) as a reference dataset. Summary-level data was merged with HapMap3 and combined association statistics of intersecting SNPs between ALS and epilepsy (including focal and generalized subtypes) with 0.05 ≤ MAF ≤ 0.5 were regressed against LD scores. **B)** Single-disorder MAF-binned GRMs were calculated to estimate liability-scale SNP-based heritability in unrelated individuals including the first 10 genetic PCs as covariates. The proportion of heritability on the liability scale stratified on MAF is shown for both diseases (y-axis) for each MAF bin (x-axis). All error bars indicate 95% confidence intervals.

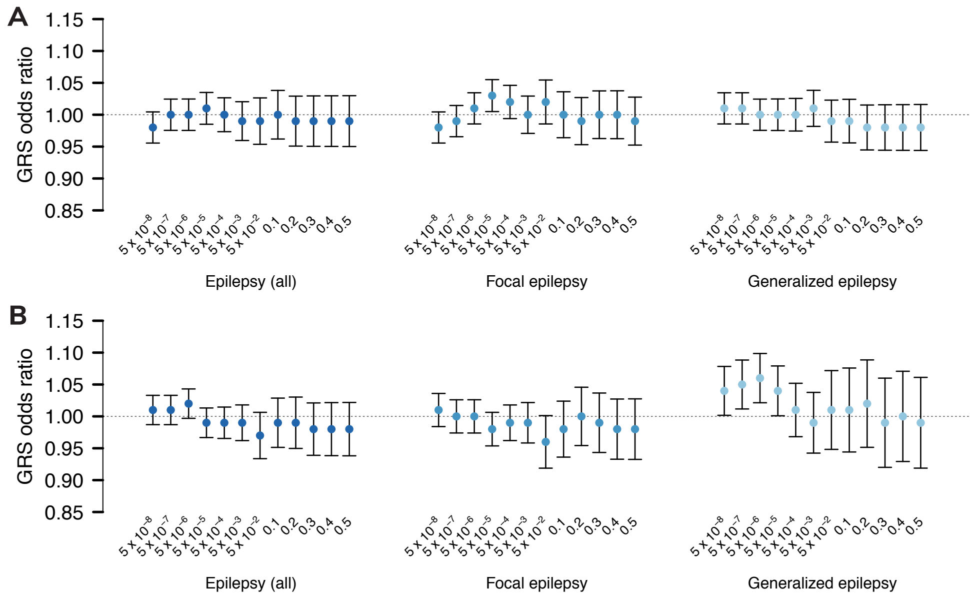


Figure 2. Genomic risk score analysis results. Twelve genomic risk scores (GRSs) were calculated per individual in the target dataset based on increasing SNP association p-values thresholds (P_T) in the discovery GWAS ($p < 5 \times 10^{-8}$ … $p < 5 \times 10^{-2}$ and $p < 0.1$ … $p < 0.5$) (x-axis). **A)** Epilepsy (including subtypes) GWAS results were used as discovery datasets, ALS was used as the target dataset. **B)** ALS GWAS results were used as the discovery dataset, epilepsy (and subtypes) were used as target datasets. Binomial regression, including sex and significant genetic principal components (PCs) (phenotype association $p < 0.05/100 = 0.0005$) as covariates, was applied to test for association between GRSs and disease phenotypes and calculate GRS effect estimates. Odds ratios (y-axis) reflect the amount by which the odds of ALS (in A) or epilepsy (in B) changes per SD increase in GRS. The p-value threshold for significant GRS effect was Bonferroni-corrected for six analyses times twelve P_T ($p < 6.94 \times 10^{-4}$). Error bars indicate 95% confidence intervals. Full regression results are shown in Supplementary Table 2.

p-value in meta-analysis ($p = 3.2 \times 10^{-8}$, OR = 1.02, similar effect directions). We looked this variant up using the available summary statistics from a recent larger ALS GWAS using the current dataset extended with additional samples [15] and found that the level of statistical association in ALS was reduced ($p = 0.11$). Sign-independent meta-analysis of ALS and generalized epilepsy strengthened association at a genome-wide significant epilepsy locus on chromosome 6q22.33 (rs13219424, ALS $p = 7.32 \times 10^{-3}$, epilepsy $p = 7.9 \times 10^{-9}$, meta $p = 3.5 \times 10^{-10}$) (Figure 3F, Supplementary Table 3B), but the statistical association of the top variant with ALS was also reduced in the larger ALS GWAS meta-analysis ($p = 0.19$) [15]. For all other loci, the level of association in either ALS or epilepsy alone was higher compared to the combined analysis.

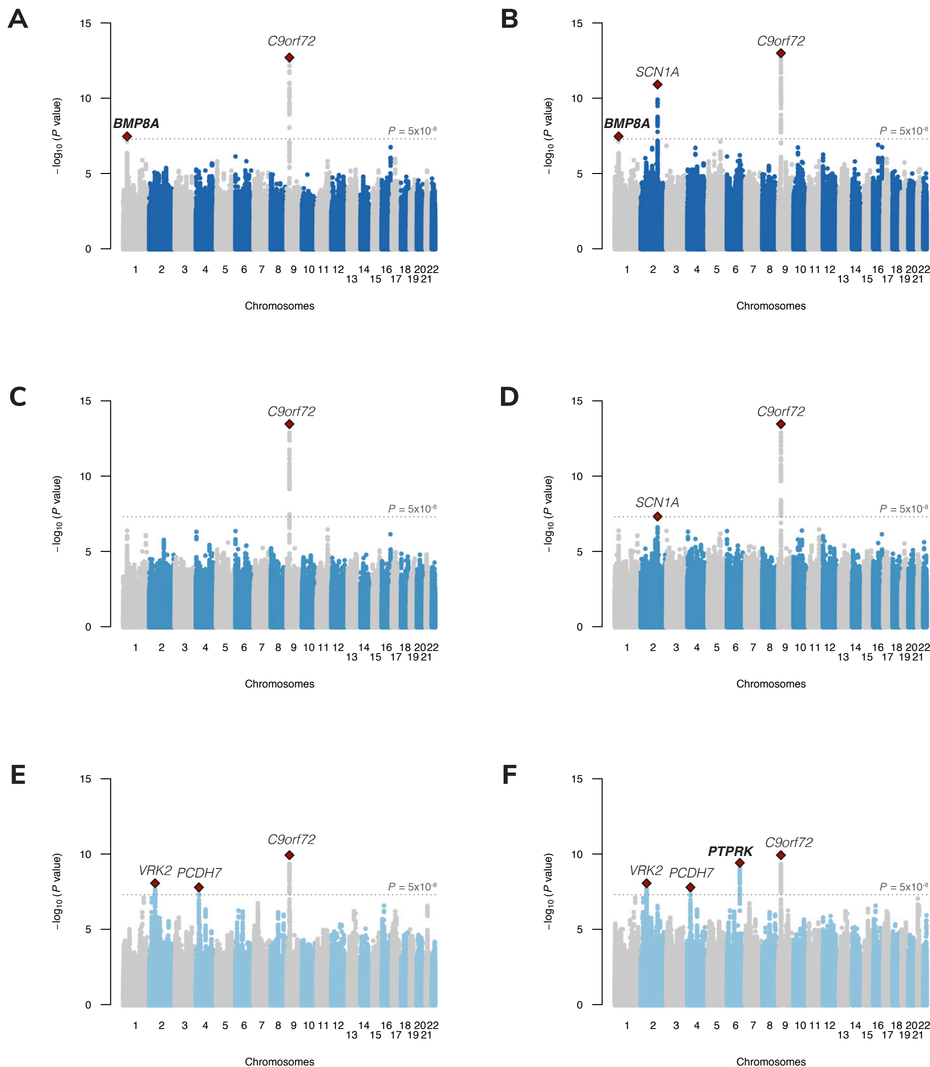


Figure 3. SNP association results of ALS-epilepsy meta-analyses. Manhattan plots show association -log_{10}-converted p-values plotted for meta-analyzed SNPs (y-axis) against their relative position on the genome (x-axis). Names of nearest genes were used to label genome-wide significant loci, those marked in bold indicate loci meeting criteria for a pleiotropic signal. Other loci are ALS- or epilepsy-specific and had decreased statistical association compared to the single-phenotype GWASs. **A)** Result for ALS-epilepsy (41,228 controls; 26,634 cases) with a pleiotropic genome-wide significant association (rs61779331, p = 3.2×10^{-8}, OR = 1.02) near the gene *BMP8A*. **B)** Sign-independent meta-analysis of ALS-epilepsy where the same pleiotropic association as in (A) was discovered. **C)** Result for ALS-focal epilepsy (41,228 controls; 21,598 cases). **D)** Sign-independent meta-analysis of ALS-focal epilepsy. **E)** Result for ALS-generalized epilepsy (41,228 controls; 15,808 cases). **F)** Sign-independent meta-analysis of ALS-generalized epilepsy with a pleiotropic association (opposing effect directions between ALS and epilepsy, rs13219424, p = 3.5×10^{-10}, OR = 1.02) near the gene *PTPRK*. See Supplementary Table 3 for detailed statistics of lead SNPs.

7.4 Discussion

Here, we used the latest and largest individual level genotype datasets for ALS and epilepsy to perform the most comprehensive cross-disorder genome-wide study between these diseases available to date. Most importantly, this enabled careful control of sample overlap and relatedness between datasets. Also, many tools that allow for cross-disorder analyses have increased power when using individual level genotype data instead of summary statistics. Overall, we showed that genetic correlations between ALS and epilepsy – including separate estimates for focal and generalized subtypes – are not significantly different from zero, that genetic architectures differ, that genomic risk scores (GRSs) capturing combined effects of common variants in one disease do not explain phenotypic variance in the other disease and that no credible loci with an effect in both ALS and epilepsy could be identified in meta-analyses.

Although all r_g estimates were non-significant, a high positive correlation for ALS and focal epilepsy resulted from LDSC ($r_g = 0.49$) which seems discrepant from the bivariate GREML estimate ($r_g = 0.052$). However, a comparison of GREML and LDSC for the estimation of r_g has shown higher accuracy and less bias in GREML estimates [16]. The inability to explain phenotypic variance in both diseases using GRSs is an expected consequence of the lack of polygenic correlation and supports the results of the GREML analyses.

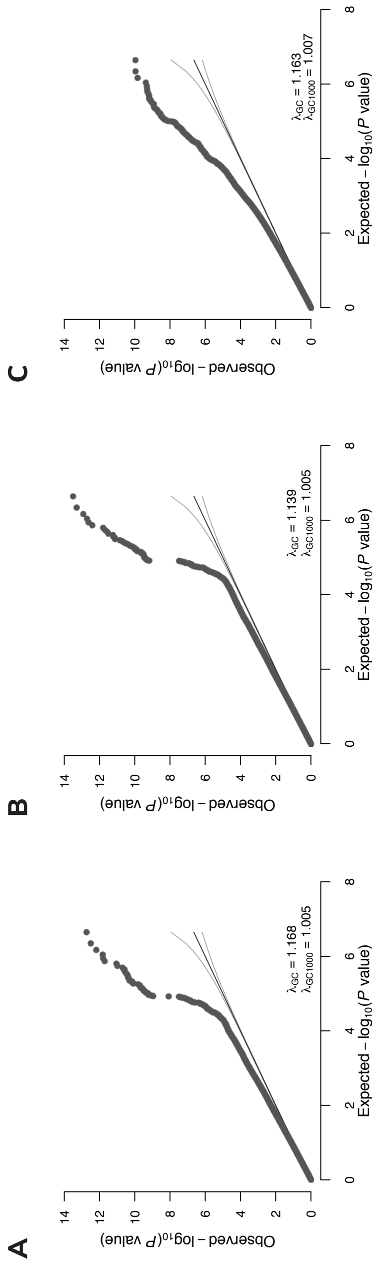
Only the genome-wide significant locus near *BMP8A* in the ALS-epilepsy meta-analysis could be considered as a true pleiotropic locus, because of its nominal significant association in ALS and epilepsy separate, corresponding direction of effect and increased statistical association in meta-analysis. The locus near *PTPRK* was pleiotropic when effect directions were ignored. Nevertheless, the top SNPs of these loci were weakly associated with ALS and did not even reach nominal significance in a more recent and larger ALS GWAS [15]. Since only summary statistics of that study were available at the time of the conduction of this study, we were unable to perform all our analyses using those data. Moreover, the contribution of epilepsy to these signals was large with association p-values around the genome-wide significance threshold [6]. These results do not support the notion that these loci are associated with both diseases.

Our results find support in epidemiological data showing a complete absence of comorbidity between ALS and epilepsy [17]. Although common and rare variants in genes coding for sodium and potassium channels have been associated with epilepsy, and an involvement of these channels in the development of cramps in ALS through deregulated excitability has been proposed [1,6,18], our results suggest that these clinical observations do not share an underlying polygenic mechanism. Cramps in ALS might thus be a consequence of disease-specific pathology leading to a hyper-excitable state of motor neurons. Nonetheless, anti-epileptic drugs have proven efficacy in ALS, albeit merely symptomatic [4].

Our findings should be regarded in light of some limitations. First, it cannot be ruled out that in even larger studies correlation estimates could get more accurate and possibly significant. However, such a correlation is then expected to be very small as we have shown no sign of substantial shared common genetic risk with the current large GWAS datasets. Second, rare genetic variation plays a role in both diseases. While we found non-significant r_g estimates in the lower MAF spectrum of our data, variants with MAF < 0.01 cannot be reliably included in these GWAS analyses. Future GWAS studies employing rare-variant imputation or large-scale sequencing efforts can provide more insight into shared heritability in the rare variant spectrum.

7.5 Supplementary information

7.5.1 Supplementary figures



Supplementary Figure 1. Observed versus expected p-values in ALS-epilepsy meta-analyses. Quantile-quantile plots for meta-analyses of **A)** ALS-all epilepsy, **B)** ALS-focal epilepsy, and **C)** ALS-generalized epilepsy. Inflation factor lambda λ_{GC} (and lambda scaled for a study with 1,000 cases and 1,000 controls, λ_{GC1000}) are shown in each plot.

7.5.2 Supplementary tables

Supplementary Table 1. Genomic regions with high LD removed prior to polygenic risk score calculation.

Chromosome	Basepair start	Basepair end
1	48000000	52000000
2	86000000	100500000
	183000000	190000000
3	47500000	50000000
	83500000	87000000
5	44500000	50500000
	129000000	132000000
6	25500000	33500000
	57000000	64000000
	140000000	142500000
7	55000000	66000000
8	8000000	12000000
	43000000	50000000
	8135000	12000000
	112000000	115000000
10	37000000	43000000
11	87500000	90500000
12	33000000	40000000
17	40900000	45000000
20	32000000	34500000

Positions refer to human genome build GRCh37

Supplementary Table 2. Genomic risk score analysis results.

A															
	Discovery dataset														
	All Epilepsy					Focal Epilepsy					Generalized Epilepsy				
P_T	SNPs	OR	se	Delta R^2	p	SNPs	OR	se	Delta R^2	p	SNPs	OR	se	Delta R^2	p
5×10^{-8}	4	0.98	1.2×10^{-2}	1.4×10^{-4}	0.06	1	0.98	1.2×10^{-2}	6.7×10^{-5}	0.20	8	1.01	1.2×10^{-2}	1.1×10^{-5}	0.60
5×10^{-7}	16	1.00	1.3×10^{-2}	1.1×10^{-6}	0.87	3	0.99	1.2×10^{-2}	2.5×10^{-5}	0.43	20	1.01	1.2×10^{-2}	3.3×10^{-5}	0.37
5×10^{-6}	55	1.00	1.3×10^{-2}	1.8×10^{-6}	0.83	30	1.01	1.2×10^{-2}	3.8×10^{-5}	0.33	46	1.00	1.2×10^{-2}	2.6×10^{-6}	0.80
5×10^{-5}	206	1.01	1.3×10^{-2}	1.1×10^{-5}	0.60	156	1.03	1.3×10^{-2}	2.4×10^{-4}	0.02	180	1.00	1.3×10^{-2}	2.4×10^{-6}	0.81
5×10^{-4}	1,092	1.00	1.4×10^{-2}	3.1×10^{-7}	0.93	847	1.02	1.3×10^{-2}	7.1×10^{-5}	0.19	799	1.00	1.3×10^{-2}	5.3×10^{-6}	0.72
5×10^{-3}	5,244	0.99	1.6×10^{-2}	1.3×10^{-5}	0.57	4517	1.00	1.5×10^{-2}	4.4×10^{-6}	0.74	4225	1.01	1.4×10^{-2}	2.6×10^{-5}	0.42
5×10^{-2}	24,620	0.99	1.9×10^{-2}	3.7×10^{-6}	0.76	22507	1.02	1.8×10^{-2}	4.5×10^{-5}	0.29	21052	0.99	1.7×10^{-2}	6.0×10^{-6}	0.70
0.1	38,975	1.00	1.9×10^{-2}	6.3×10^{-7}	0.90	36594	1.00	1.8×10^{-2}	5.7×10^{-7}	0.91	34571	0.99	1.7×10^{-2}	3.4×10^{-6}	0.77
0.2	58,330	0.99	2.0×10^{-2}	1.1×10^{-5}	0.61	55623	0.99	1.9×10^{-2}	7.9×10^{-6}	0.66	53535	0.98	1.8×10^{-2}	3.3×10^{-5}	0.37
0.3	72,518	0.99	2.0×10^{-2}	7.7×10^{-6}	0.66	69907	1.00	1.9×10^{-2}	1.6×10^{-6}	0.84	68111	0.98	1.8×10^{-2}	4.8×10^{-5}	0.28
0.4	83,813	0.99	2.0×10^{-2}	6.9×10^{-6}	0.68	81115	1.00	1.9×10^{-2}	3.7×10^{-7}	0.92	79566	0.98	1.8×10^{-2}	4.1×10^{-5}	0.32
0.5	92,480	0.99	2.0×10^{-2}	8.8×10^{-6}	0.64	89911	0.99	1.9×10^{-2}	2.9×10^{-6}	0.79	88794	0.98	1.8×10^{-2}	3.0×10^{-5}	0.39

< Supplementary Table 2 continued.

B

		Target dataset											
		All Epilepsy				Focal Epilepsy				Generalized Epilepsy			
P_T	SNPs	OR	se	Delta R^2	p	OR	se	Delta R^2	p	OR	se	Delta R^2	p
5×10^{-8}	4	1.01	1.2×10^{-2}	5.6×10^{-5}	0.23	1.01	1.3×10^{-2}	7.7×10^{-6}	0.69	1.04	2.0×10^{-2}	2.9×10^{-2}	0.05
5×10^{-7}	8	1.01	1.2×10^{-2}	1.3×10^{-5}	0.57	1.00	1.3×10^{-2}	5.2×10^{-6}	0.74	1.05	2.0×10^{-2}	4.2×10^{-2}	0.02
5×10^{-6}	17	1.02	1.2×10^{-2}	6.6×10^{-5}	0.19	1.00	1.3×10^{-2}	1.0×10^{-6}	0.89	1.06	2.0×10^{-2}	7.2×10^{-2}	0.002
5×10^{-5}	86	0.99	1.2×10^{-2}	7.6×10^{-6}	0.66	0.98	1.3×10^{-2}	8.4×10^{-5}	0.19	1.04	2.0×10^{-2}	2.9×10^{-2}	0.05
5×10^{-4}	483	0.99	1.3×10^{-2}	1.6×10^{-5}	0.53	0.99	1.4×10^{-2}	2.6×10^{-5}	0.46	1.01	2.1×10^{-2}	5.4×10^{-4}	0.79
5×10^{-3}	2,836	0.99	1.4×10^{-2}	1.4×10^{-5}	0.56	0.99	1.6×10^{-2}	3.2×10^{-5}	0.42	0.99	2.4×10^{-2}	7.5×10^{-4}	0.75
5×10^{-2}	17,125	0.97	1.9×10^{-2}	7.9×10^{-5}	0.16	0.96	2.1×10^{-2}	1.6×10^{-4}	0.07	1.01	3.2×10^{-2}	7.0×10^{-4}	0.76
0.1	28,319	0.99	2.0×10^{-2}	2.1×10^{-5}	0.46	0.98	2.2×10^{-2}	4.3×10^{-5}	0.35	1.01	3.4×10^{-2}	5.7×10^{-4}	0.78
0.2	46,467	0.99	2.1×10^{-2}	2.8×10^{-6}	0.79	1.00	2.3×10^{-2}	1.7×10^{-6}	0.85	1.02	3.5×10^{-2}	1.9×10^{-3}	0.62
0.3	60,997	0.98	2.1×10^{-2}	2.9×10^{-5}	0.39	0.99	2.4×10^{-2}	1.0×10^{-5}	0.65	0.99	3.6×10^{-2}	2.7×10^{-4}	0.85
0.4	73,344	0.98	2.1×10^{-2}	3.7×10^{-5}	0.33	0.98	2.4×10^{-2}	2.5×10^{-5}	0.47	1.00	3.6×10^{-2}	3.0×10^{-5}	0.95
0.5	83,607	0.98	2.1×10^{-2}	4.4×10^{-5}	0.29	0.98	2.4×10^{-2}	3.5×10^{-5}	0.40	0.99	3.6×10^{-2}	1.9×10^{-4}	0.88

A) Results using epilepsy (including subtypes) as discovery datasets and ALS as a target dataset. Explained variance in ALS (case-control status) by epilepsy GRSs was tested using binomial regression in a generalized linear model, including sex and significant principal components ($p < 0.05/100 = 5x10^{-4}$; PCs 1, 2, 3, 5, 12, 17, 18, 19, 21, 26, 28, 29, 33, 43, 45, 46, 47, 53, 65, 70, 75, 96) as covariates.

B) Results using ALS as a discovery dataset and epilepsy (including subtypes) as target datasets. Explained variance in epilepsy (case-control status) by ALS GRSs was tested using binomial regression in a generalized linear model including sex and significant principal components ($p < 0.05/100 = 5x10^{-4}$; all epilepsy: PCs 1, 2, 3, 6, 7, 22, 24, 31, 95; focal epilepsy: PCs 1, 2, 3, 6, 7, 8, 22, 24, 31; generalized epilepsy: PCs1, 2, 3, 7, 8, 9, 24, 44, 66, 95) as covariates.

GRSs were scaled around mean 0 with variance 1. OR reflects the odds for disease per one unit (one SD) increase of the GRS at the respective P_T.

Table columns: P_T, p-value threshold of SNP association in discovery data; SNPs, number of SNPs in P_T (in B this is the maximum number of SNPs due to missing SNP genotypes; OR, odds ratio of the GRS; se, standard error of the effect estimate of the GRS; Delta R^2, difference in explained variance (Nagelkerke R^2) compared to baseline model without GRS; p, p-value of the GRS in the model.

Results in bold show GRS effects with $p < 0.05$, although none survived multiple-testing correction (Bonferroni-correction for four traits times twelve P_T included, $p < 0.001$).

Supplementary Table 3. Meta-analysis association results of lead SNPs.

A. Normal meta-analysis													
				ALS			Epilepsy			Meta-analysis			
Chr	bp	SNP	Risk allele	OR	se	p	OR	se	p	OR	se	p	Direction
ALS-All epilepsy													
1	39970928	rs61779331	A	1.01	4.6×10^{-3}	2.3×10^{-2}	1.02	4.0×10^{-3}	8.7×10^{-8}	1.02	3.0×10^{-3}	3.2×10^{-8}	++
9	27545467	rs812858	T	1.05	4.6×10^{-3}	1.2×10^{-24}	1.00	4.1×10^{-3}	4.7×10^{-1}	1.02	3.1×10^{-3}	1.8×10^{-13}	++
ALS-Focal epilepsy													
9	27545467	rs812858	T	1.05	4.6×10^{-3}	1.2×10^{-24}	1.00	4.1×10^{-3}	3.1×10^{-1}	1.02	3.1×10^{-3}	3.2×10^{-14}	++
ALS-Generalized epilepsy													
2	58042241	rs1402398	G	1.00	4.0×10^{-3}	0.54	1.02	2.7×10^{-3}	6.0×10^{-11}	1.01	2.3×10^{-3}	8.0×10^{-9}	++
4	31148846	rs1463849	G	1.00	4.0×10^{-3}	0.27	1.02	2.7×10^{-3}	9.3×10^{-10}	1.01	2.3×10^{-3}	1.5×10^{-8}	++
9	27545960	rs700791	A	1.05	4.6×10^{-3}	1.2×10^{-24}	1.00	3.3×10^{-3}	5.2×10^{-1}	1.02	2.7×10^{-3}	1.1×10^{-10}	++

Table columns: Chr: chromosome; bp: basepair position (hg19/GRCh37); SNP: variant rs-number; Risk allele: disease-associated allele; OR: odds ratio of SNP in ALS, epilepsy and meta-analysis; se: standard error of SNP effect estimate in ALS, epilepsy and meta-analysis; p: association p-value of SNP in ALS, epilepsy and meta-analysis; Direction: direction of effect in ALS and epilepsy.

< Supplementary Table 3 continued.

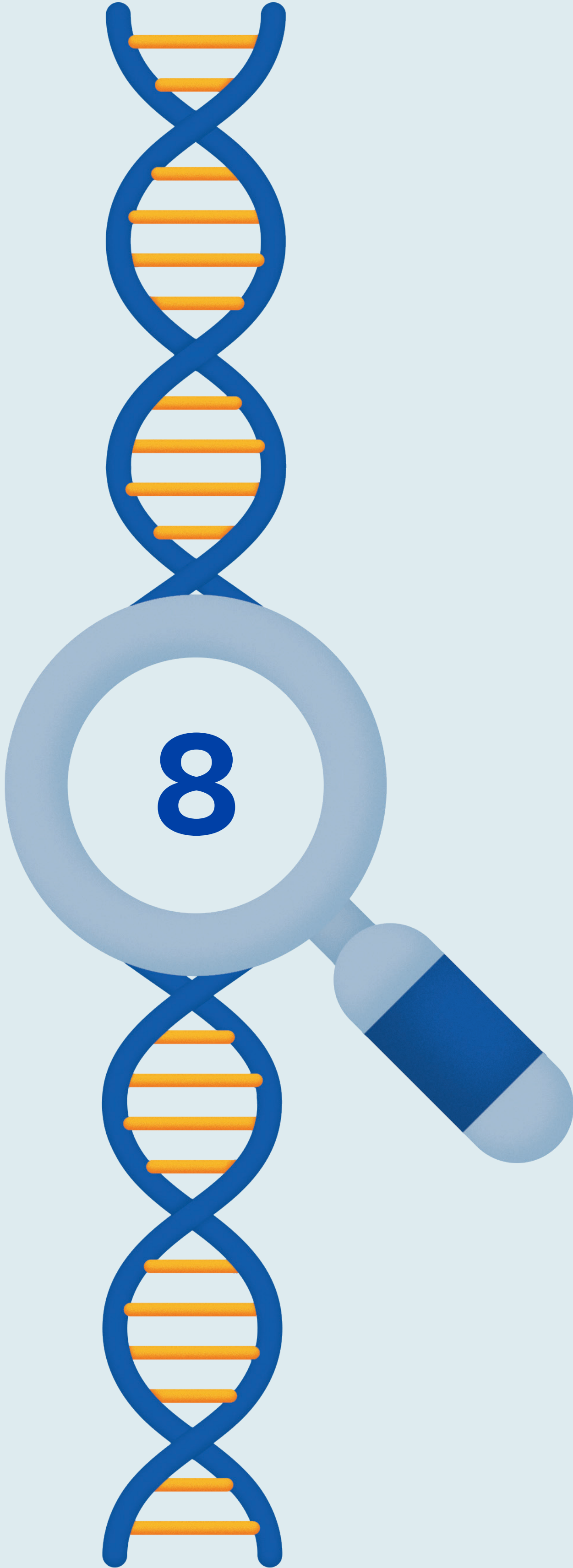
B. Sign-independent meta-analysis

Chr	bp	SNP	Alleles	p ALS	p Epilepsy	Meta-analysis OR	se	p
ALS-All epilepsy								
1	39970928	rs61779331	A/C	2.3×10^{-2}	8.7×10^{-8}	1.02	3.00×10^{-3}	3.19×10^{-8}
2	166998767	rs6432877	C/G	8.0×10^{-2}	8.9×10^{-14}	1.02	3.00×10^{-3}	1.14×10^{-11}
9	27563755	rs2484319	A/C	1.9×10^{-26}	0.57	1.02	3.00×10^{-3}	9.40×10^{-14}
ALS-Focal epilepsy								
2	166998767	rs6432877	C/G	8.0×10^{-2}	8.9×10^{-14}	1.02	3.00×10^{-3}	4.46×10^{-8}
9	27545467	rs812858	C/T	1.2×10^{-24}	0.31	1.02	3.10×10^{-3}	3.15×10^{-14}
ALS-Generalized epilepsy								
2	58042241	rs1402398	A/G	0.54	6.0×10^{-11}	1.01	2.30×10^{-3}	7.94×10^{-9}
4	31148846	rs1463849	A/G	0.27	9.3×10^{-10}	1.01	2.30×10^{-3}	1.48×10^{-8}
6	128332084	rs13219424	T/C	7.3×10^{-3}	7.9×10^{-9}	1.02	2.40×10^{-3}	3.50×10^{-10}
9	27497988	rs10967965	A/T	8.3×10^{-22}	0.28	1.02	2.70×10^{-3}	1.08×10^{-10}

Table columns: Chr: chromosome; bp: basepair position (hg19/GRCh37); SNP: variant rs-number; Alleles: alleles of the SNP; OR: odds ratio of SNP meta-analysis; se: standard error of SNP effect estimate in meta-analysis; p: association p-value of SNP in ALS, epilepsy and meta-analysis.

7.6 References

1. Kanai K, Kuwabara S, Misawa S, Tamura N, Ogawara K, Nakata M, et al. Altered axonal excitability properties in amyotrophic lateral sclerosis: impaired potassium channel function related to disease stage. Brain 2006;129(4):953–62.
2. Kanai K, Shibuya K, Sato Y, Misawa S, Nasu S, Sekiguchi Y, et al. Motor axonal excitability properties are strong predictors for survival in amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 2012;83(7):734–8.
3. Shimizu T, Bokuda K, Kimura H, Kamiyama T, Nakayama Y, Kawata A, et al. Sensory cortex hyperexcitability predicts short survival in amyotrophic lateral sclerosis. Neurology 2018;90(18):e1578–87.
4. Kovalchuk MO, Heuberger JAAC, Sleutjes BTHM, Ziagkos D, van den Berg LH, Ferguson TA, et al. Acute Effects of Riluzole and Retigabine on Axonal Excitability in Patients With Amyotrophic Lateral Sclerosis: A Randomized, Double-Blind, Placebo-Controlled, Crossover Trial. Clin Pharmacol Ther 2018;9(Pt 3):65.
5. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet 2016;48(9):1043–8.
6. The International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun 2018;9(1):5269.
7. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4(1):7.
8. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet 2011;88(1):76–82.
9. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.
10. Lee SH, DeCandia TR, Ripke S, Yang J, Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ), International Schizophrenia Consortium (ISC), et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 2012;44(3):247–50.
11. Johnston CA, Stanton BR, Turner MR, Gray R, Blunt AH-M, Butt D, et al. Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. J Neurol 2006;253(12):1642–3.
12. Semah F, Picot MC, Adam C, Broglin D, Arzimanoglou A, Bazin B, et al. Is the underlying cause of epilepsy a major prognostic factor for recurrence? Neurology 1998;51(5):1256–62.
13. Speed D, O'Brien TJ, Palotie A, Shkura K, Marson AG, Balding DJ, et al. Describing the genetic architecture of epilepsy through heritability analysis. Brain 2014;137(Pt 10):2680–9.
14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190–1.
15. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron 2018;97(6):1268–1283.e6.
16. Ni G, Moser G, Psychiatric GWAS Consortium Bipolar Disorder Working Group, Wray NR, Lee SH. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. Am J Hum Genet 2018;102(6):1185–94.
17. Tartaglia MC, Rowe A, Findlater K, Orange JB, Grace G, Strong MJ. Differentiation Between Primary Lateral Sclerosis and Amyotrophic Lateral Sclerosis. Arch Neurol 2007;64(2):232–6.
18. Helbig I, Scheffer IE, Mulley JC, Berkovic SF. Navigating the channels and beyond: unravelling the genetics of the epilepsies. Lancet Neurol 2008;7(3):231–45.


8

Chapter 8

Summary and general discussion

This thesis contains a number of genome-wide studies in complex psychiatric and neurological diseases with the overarching goal of improving our understanding of the etiology of these diseases, the biological mechanisms that are possibly at play and the genetic landscape in which these diseases are present. ***PART 1*** covered single-disorder studies where we investigated the enrichment of common genetic variants associated with schizophrenia in biologically annotated gene sets (***chapter 2***), performed genome-wide association analyses on biochemical and behavioral phenotypes related to traumatic stress and the development of psychopathology (***chapter 3***), and combined the effects of common variants associated to PTSD and depression to predict the development of symptoms related to these diseases after military deployment (***chapter 4***). In ***PART 2***, cross-disorder studies were presented where common approaches and their clinical utility in psychiatry were outlined (***chapter 5***), and shared common genetic risk between ALS and schizophrenia (***chapter 6***) and ALS and epilepsy (***chapter 7***) were examined using large GWAS datasets. The current chapter summarizes the main findings from these chapters and adds further discussion.

8.1 Insight into biological mechanisms of disease through gene set enrichment analysis

Gene set enrichment analysis (GSEA) using association results of common genetic variants can provide a first step in the elucidation of disease mechanisms by collapsing the combined effects of millions of variants tested in GWAS into a lower number of genes and an even more limited number of gene sets, advancing the interpretability of GWAS results. The development of novel statistical methods to perform GSEA has allowed to generate novel – and validate existing – hypotheses about the biological pathways relevant to (psychiatric) disease.

8.1.1 Enrichment of schizophrenia-associated variants in post-synaptic signaling

For schizophrenia, the understanding of the disease's pathophysiology is still very limited and the development of successful treatments has not seen a major breakthrough since the introduction of antipsychotic mediation in the 1960s and 1970s [1]. Nonetheless, the disease can be considered as one of the success

stories in GWAS of brain diseases as the number of identified disease risk loci has greatly expanded with the increase in sample size [2-4] in well-powered studies that have shown a clear role for SNPs in the etiology of the disease ($h_S^2 = 0.23$) [5]. The interpretation of schizophrenia GWAS findings becomes however more challenging with the discovery of additional loci. To that end, ***chapter 2*** used a comprehensive GSEA on recent results from large schizophrenia GWASs [6].

Starting with a hypothesis-free approach, enrichment of common genetic variants associated with schizophrenia in gene sets related to the regulation of neuron differentiation and synaptic plasticity were found. This was followed up by analyses focused on synaptic signaling using molecular-level pathways of the five major neurotransmitter systems (dopamine, acetylcholine, gamma-aminobutyric acid, glutamate and serotonin), long-term potentiation and long-term depression. Significant enrichment in the dopaminergic and cholinergic synapses and long-term potentiation was found. Enriched genes and molecules in the synapse were prioritized and insight into this network of synaptic signaling molecules was provided, highlighting strong association of genes that are represented in the synapse of multiple neurotransmitter systems (e.g. *CACNA1C*) and system-specific genes (e.g. *DRD2*). Overall, the biological systems found to be enriched agreed with previous GSEA studies in smaller schizophrenia GWAS datasets with regard to the involvement of synaptic signaling [7-9] and possible role for calcium signaling [9,10]. In contrast, no significant enrichment in immune-related biological processes was found while these have been implicated in earlier studies [9,11]. This finding was somewhat surprising given the strong association of the major histocompatibility complex (MHC) at population-level, found through GWASs [3,4], and the evidence for the role of microglia in inflammatory processes in schizophrenia [12]. Of interest, MHC molecules are proposed modulators of plasticity in neuronal synapses during critical periods of development [13]. Moreover, common structural variation in complement component 4 (C4) genes – of which the protein products localize to synapses, dendrites, axons and cell bodies in the brain – has been hypothesized to play a role in neurodevelopment and the reduced number of synapses in the brains of schizophrenia patients [14]. These results point towards a close association between immune- and synaptic processes in the development of schizophrenia. The lack of enrichment in immune-related

biological processes in our study might have a methodological origin, as further elaborated upon in the next paragraph.

8.1.2 Methodological considerations in gene set enrichment analysis

Although GSEA for schizophrenia broadly point towards the same underlying mechanisms, variation in findings from GSEA studies can be explained by differences in 1) tested gene sets, 2) the applied enrichment analysis method, and 3) GWAS results used as input. We used unstructured gene sets from gene ontology (GO), included in the Molecular Signatures Database (MSigDB) [15], in our primary analysis. In the follow-up analyses, we used well-characterized biochemical pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) which merely represented synaptic signaling pathways [16]. Although these gene sets reflect relevant biological systems and were derived from curated and regularly updated databases, differences in the composition of gene sets reflecting similar biological processes exist between databases, which is mainly attributable to different sources of information used to build gene sets (e.g. protein interaction data, gene co-expression or experimental data) [17]. While immune processes were part of the tested gene sets, it might be that the signal of associated genes got diluted by non-associated genes present in these gene sets or that an underrepresentation of schizophrenia-associated genes in these sets resulted in the low enrichment signal.

A multitude of GSEA tools have been released over the past decade, with similarities in underlying statistical structures but with differences in their performance and reliability [18]. As a primary method, we used Multi-marker Analysis of GenoMic Annotation (MAGMA), a relatively new GSEA tool that maps SNPs to genes, calculates gene p-values, and tests enrichment of associated genes in gene sets while comprehensively correcting for confounding factors such as gene size, gene density and linkage disequilibrium (LD) [19]. As we found no difference in the primary analysis result with or without exclusion of the MHC region, this is an indication that MAGMA indeed corrects for bias that might be induced by this region with complex LD. On the other hand, the low enrichment in biological processes including MHC genes could also be a result of MAGMA's LD correction suppressing the association signal in this region. In contrast to MAGMA, Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) [20], which we used and has been commonly used in

other GSEA studies, has a higher type-1 error rate [18]. This could be reflected in the widespread enrichment found in synaptic pathways when using the latest schizophrenia GWAS including the CLOZUK cohort [4] compared to the enrichment result of other methods. INterval enRICHment analysis (INRICH) [21], a third method applied, also performs well at the correction for confounding effects [18]. The difference in performance and reliability towards correction for type-1 error rates makes it useful to explore and compare multiple tools for GSEA. Although slight method-dependent differences in enrichment in synaptic signaling pathways were thus found in ***chapter 2***, the overall correspondence was high.

Results from the most recent and largest schizophrenia GWASs were used [3,4]. Although GSEA works best on well-powered GWASs of polygenic phenotypes, it is doubtful whether gene set analyses based on future larger schizophrenia GWASs will provide additional insights into the biological mechanisms underlying the disease. In the competitive analysis approach used in GSEA, the disease-association of genes in a gene set is tested against a background of gene associations outside the gene set. As such, this background contains a high number of associated genes in polygenic diseases like schizophrenia, ultimately reducing the contrast between gene set and background and attenuating or even a decreasing power [18]. Estimates of up to 10,000 independent SNP associations in schizophrenia [22] might complicate future GSEA.

8.1.3 Is gene set enrichment analysis outdated, or does it need to be updated?

Besides the possible limitation of future GSEA by abundant polygenic signal, the approach can be criticized because it only takes into account SNPs that map onto genes. This 'nearest gene' approach builds on the assumption that a SNP has an effect on the gene that it maps onto. Moreover, although SNPs in close proximity to genes – located in possible regulatory regions [23] – are often captured in GSEA by extending gene up- and downstream boundaries, the non-coding regions of the genome are excluded. Recent insights in the field of genetics and functional genomics have revealed that the 'nearest gene' assumption is partly obsolete and that non-coding variants can have highly relevant regulatory functions. For example, three-dimensional chromatin interaction capturing (for example through Hi-C [24]) has identified

schizophrenia-associated SNPs exerting effects on regulatory regions of linearly distant genes, but that might be close to the variant due to the folding of the chromatin [25].

In the past years, large-scale annotation efforts have resulted in genome-wide knowledge on tissue-specific expression and genotype-dependent expression (the Gene- and Tissue Expression (GTEx) database) [26,27]; the abundance of chromatin interactions, regulatory elements and promoter regions across the human genome (the Encyclopedia of DNA elements, ENCODE) [28]; epigenomic hallmarks (Roadmap Epigenome database) [29]; and databases focused on gene-drug interactions [30]. Additionally, the ability to apply genetic and genomics techniques to single cells provides even higher resolution functional insight into the role of variants associated with schizophrenia and other psychiatric diseases [31]. Although these multiple layers of information are challenging to implement into one comprehensive method for the interpretation of GWAS results, annotations from such databases and studies can refine gene sets and pathways used in GSEA and can potentially be used to prioritize and weigh variants based on their functional characteristics [17].

GSEA thus remains an important method for the interpretation of GWAS results, as presented in ***chapter 2***, whereas it can become more powerful with the implementation of novel sources of functional annotation. Prioritized biological pathways and genes important to schizophrenia can further guide translational experiments of synaptic signaling in the disease, for example using in vitro brain organoids to study the impact of these genes on connectivity and physiology of neuronal networks [32].

8.2 Variant discovery and risk prediction in stress and trauma-related psychopathology

8.2.1 Phenotypic heterogeneity in psychiatric disease and the impact on GWAS

While for most psychiatric diseases the number of discovered risk loci is ever expanding and biological insight is rapidly resulting from these studies, other psychiatric disorders are lagging behind. Out of several reasons, such as smaller sample size or possible lower involvement of a polygenic architecture, an

important point of debate is the phenotypic classification in psychiatry. GWASs are originally based on a case-control design and psychiatric patients included in these studies are most often diagnosed according to DSM-5 criteria [33]. Many of the disorders show symptomatic overlap which is reflected in low test-retest reliability and high levels of comorbidity when DSM-based diagnoses are used [34], implying that current diagnostic criteria are non-specific and borders are vague. This might impact findings from GWAS in two ways: First, in successful GWASs with many loci discovered (such as schizophrenia) the question arises as to how relevant all risk loci are at the level of the individual patient. Second, a broad clinical definition and comorbidity lead to phenotypic heterogeneity (and the risk of diagnostic misclassification), negatively impacting the power in case-control GWASs to discover risk variants [35,36]. On the one hand, this loss of power can be partially compensated by further increasing the sample size, but potential gains in power could also be achieved when the focus is shifted towards more detailed assessment of quantitative (intermediate) phenotypes underlying psychiatric conditions [37].

8.2.2 GWAS of *stress-related quantitative phenotypes in a deployed military cohort*

A range of psychiatric symptoms may arise in the aftermath of one or multiple traumatic events, in severe cases leading to diagnosis with PTSD or depression. For individuals predisposed to experience trauma, such as military personnel engaging in direct combat, the lifetime risk of PTSD is higher (13.2%) than for the general population (5.6%) [38,39]. Beyond these diagnosed cases, an even larger proportion of military personnel report symptoms of post-traumatic stress and depression after deployment [40].

GWASs specifically targeting PTSD have thus far employed case-control designs, where cases were ascertained through DSM-5-based questionnaires or clinical assessments [41-46]. These GWASs had mixed success in the discovery and replication of risk loci or the estimation of heritability attributable to common variants. Furthermore, identified loci have not replicated across studies and cohorts, which can be explained by differences in sex, genetic ancestry, type of trauma and method of case ascertainment [47]. Although PTSD is the best scoring psychiatric disease in field trials assessing inter-rater reliability (kappa value of 0.69) [34], the diagnostic reliability is still imperfect and together

with the heterogeneous presentation of the disease this could be a major limitation in the discovery of relevant loci and biological mechanisms.

The heterogeneous presentation of trauma-induced psychopathology and relatively little success of case-control GWASs in related clinical phenotypes render the exploration of alternative approaches in these studies timely. GWAS of quantitative phenotypes and post-trauma psychopathology may highlight variants with a more direct involvement in relevant biological processes. In ***chapter 3***, a multivariate analysis of quantitative biochemical and behavioral phenotypes related to stress mediation and psychiatric symptoms in a military cohort was described with the aim to discover variants more closely associated with biological processes possibly involved in the development of psychopathology. Subjects had been exposed to potential trauma during deployment (including direct combat) and both pre- and post-deployment measurements were available such that genetic variants underlying the response to traumatic stress could be targeted. One genome-wide significant locus (lead SNP rs10100651) near the gene *INTS8* on chromosome 8 was discovered and multiple phenotypes contributed to this signal, indicating a pleiotropic effect of this locus on stress- and trauma-related phenotypes.

Replication of discovered GWAS loci in an independent sample has become a standard procedure and is required to ensure that a result is a disease-specific association that was not driven by confounding factors [48]. Given the unique longitudinally collected set of phenotypes used in our study, we were unable to find an independent replication sample in which similar phenotypes were available, although several other cohorts in which stress and psychopathology are studied longitudinally in military personnel exist [49,50]. Nonetheless, a lookup strategy resulted in quasi-replication at nominal significance in a GWAS of self-reported PTSD in the UK Biobank. Although this locus may eventually be discovered in association with PTSD, the construct of phenotypes assessed in our analysis and the predominance of hormones related to the stress response in the optimal association model at this locus (DHEA-S, GABA, NPY, SHBG) could also imply a more fundamental function of this gene in modulation of the stress response, regardless of psychopathology. Through the availability of GWAS results for a large number of phenotypes, phenome-wide scans (or phenome-wide association studies, PheWAS) have nowadays become very

useful to test association of one or a few variants or genes against a multitude of traits in different phenotypic domains [51]. Thus, the more central role of the disocvered variant in stress-associated phenotypes was confirmed by a phenome-wide scan where numerous stress-related conditions – including blood pressure, cardiovascular disease, insomnia and subjective wellbeing – were strongly associated with genes surrounding this locus and functionally linked to the lead SNP (*TP53INP1* and *INTS8*, prioritized through analysis of eQTL and chromatin interactions). In addition to the phenome-wide findings, *TP53INP1* has also been associated with Alzheimer's disease (AD) [52], while accelerated ageing of neurons has been proposed in PTSD [53] and there is a higher rate of amyloid plaque accumulation and AD among veterans [54,55]. Overall, the evidence from alternative replication efforts thus supports the role of this locus and surrounding genes in neuropsychiatric phenotypes.

8.2.3 Increased power and interpretability in multivariate approaches

Multivariate GWAS has shown increased statistical power for discovery of genetic associations over the use of a univariate approach [56]. In ***chapter 3*** Software for Correlated Phenotype Analysis (SCOPA) [57] was used, which is based on a reverse regression approach where regression of a SNP genotype against multiple phenotypes is performed. As compared to other tools that apply a reverse regression approach [58], SCOPA can use dosage data to integrate imputation uncertainty in the analysis and dissect the multivariate association model using Bayesian statistics to provide an optimal set of phenotypes at a genome-wide significant variant and an effect estimate (corrected for possible phenotypic correlation) for each phenotype in the model [57]. This allows to interpret the contribution that each phenotype makes to the association and increases the interpretability of a multivariate model with quantitative phenotypes over that of associations found in case-control studies. Even the relatively low sample size for GWAS standards in the study in ***chapter 3*** (~300-500 as compared to thousands in most case-control analyses) surprisingly identified a credible candidate locus possibly involved in stress regulation and trauma-induced psychopathology, as supported by the aforementioned replication efforts. When multiple (correlated) phenotypes are associated at a locus, we confirm that multivariate methods may thus leverage power for detection over univariate approaches.

As sample sizes for case-control GWASs of stress-related disorders are being increased, the field also shifts towards the use of quantitative phenotypes related to post-trauma psychopathology [47,59]. These studies are likely to reach larger sample sizes than ours, so that additional discoveries and relevant biological insights will likely result from these novel approaches in the coming years. Furthermore, future case-control GWASs of PTSD can be analyzed in combination with other psychiatric diseases, as PTSD is genetically correlated to schizophrenia and depression [44], and has a genetic overlap with bipolar disorder [42]. A recent GWAS of PTSD has also shown that there is a heritability gap between men (h_S^2 = 0.07) and women (h_S^2 = 0.27) [44]. Women are at increased risk of developing PTSD, and sex differences in the response to trauma and the development of psychopathology have been reported [60]. In our study we only included men and because the discovered association was largely driven by (gonadal) hormone levels, the finding might not be generalizable to women. Sex-specific analyses are therefore an important future point of focus in stress-related genetics research. Finally, the importance of environmental factors in mediating the effects of these variants should be further investigated in SNP x environment GWASs of stress, as illustrated by the effect of risk SNPs for depression being strongly modulated by the environment [61].

8.2.4 Prediction of psychopathology in trauma-exposed subjects

Established and novel GWAS approaches will further unravel variants and disease mechanisms involved in trauma-related psychopathology. Alternatively, early identification of individuals susceptible to develop mental health issues upon exposure to stress or trauma can result in a better recognition of symptoms and adequate intervention. Because this susceptibility is heritable, genetic factors have the potential to serve as a stable predictor. As the effect sizes of individual SNPs are extremely small, combining the effects that multiple SNPs exert on disease liability is necessary to achieve predictive power. Polygenic risk scores (PRSs), capturing the small effects that SNPs confer on disease risk into a single quantitative measurement, are extremely useful when building prediction models for heritable diseases [62,63].

Published GWASs of PTSD and depression were used to predict deployment-induced psychopathology in ***chapter 4***. Furthermore, the interaction of genetic

vulnerability with childhood and deployment trauma was investigated. Combining results of recent PTSD GWASs through meta-analysis [43,44] and the use of the most recent and largest depression GWAS [64] allowed us to calculate PRSs for these diseases in individuals in military personnel and evaluate the predictive value of these PRSs (also in interaction with trauma) on the development of psychopathology at multiple time points up to five years after deployment. This longitudinal context is relevant, as it has for example been shown that there are early onset (directly after deployment) and late onset (approximately 2 years after deployment) trajectories for PTSD [65]. Despite the use of these large discovery samples, no predictive effects of PRSs on post-deployment psychopathology were found, neither in interaction with trauma exposure during childhood or during deployment.

There are several possible reasons for the non-significant findings in this study. First, this has been one of the first studies using PRSs based on PTSD GWAS results. It is reasonable to argue that for PTSD the current GWAS results are underpowered and contain too little disease-specific signal to be suitable for PRS calculation [63]. For MDD however, the sample size was much larger resulting in PRSs with higher predictive accuracy as shown in a recent study successfully predicting the development of depression using the same discovery data for PRS calculation [66]. PRSs based on smaller MDD discovery datasets have also shown predictive effects [67]. Second, etiological differences between the case-control phenotypes in the discovery data and the quantitative (questionnaire-based) measure can play a role. For example, deployment-related depression could be mechanistically different from MDD in the general population, diluting the predictive value of such scores in a military cohort. Third, and perhaps the most likely explanation, is the absence of a strong increase in post-traumatic stress and depressive symptoms throughout deployment.

The predictive accuracy of PRSs can be increased in future studies when the phenotypic heterogeneity among cases included in the discovery GWAS is reduced, allowing computation of more predictive PRSs for specific subgroups of cases or subtypes of disease [68]. The interaction of genetic risk with trauma exposure is furthermore highly relevant, as demonstrated by other studies where polygenic risk for MDD interacted with childhood trauma to predict

depression [67], and smoking PRSs interacted with trauma exposure to predict smoking behavior [69]. As environmental factors also play a role in the development of post-deployment psychopathology, a combination of these and PRSs might ultimately aid in the identification and prevention of the development of these disabling psychiatric conditions in at-risk individuals.

8.3 Cross-disorder analyses in psychiatric and neurological disease

The first part of the thesis focused on single-disorder genetic studies, although ***chapter 3*** already touched upon pleiotropy with the discovery of a genome-wide significant locus associated with multiple (correlated) quantitative phenotypes. The insight into the polygenic architecture of psychiatric and neurological diseases in the past decade of GWAS has simultaneously shown that many variants and genes influence multiple phenotypes, suggesting biological similarities. This has opened up possibilities for genetic cross-disorder analyses, which allow to combine datasets where different traits have been measured in different individuals [70]. Relatively large study populations achieved by combining datasets have increased the power to detect pleiotropic variants associated with the risk for two or more diseases [71]. Furthermore, the combined information of millions of SNPs can be used to asses shared heritability, and thus overlapping genetic risk between disorders, through the estimation of genetic correlation [5,71-73].

8.3.1 The implications of cross-disorder studies on the field of psychiatry and neurology

In ***chapter 5*** we argued that three domains relevant to clinical practice in psychiatry may benefit from cross-disorder studies: diagnostics, prediction and treatment. The concepts explained in that chapter are generalizable to other clinical fields, such as neurology. Psychiatry is however exceptional in the sense that no other phenotypic category, except perhaps anthropometric or metabolic traits, has shown such a high rate of genetic correlations [72].

The field of psychiatry studies mental disorders without visible pathological hallmarks with a focus on symptomatology as experienced by the patient or observed by the clinician [74]. This approach complicates accurate disease classification based on etiological similarities, partly reflected in the high

genetic correlations, and has led to debate about possible improvements to this system. Insights from cross-disorder studies in psychiatric genetics can help reshape the diagnostic landscape by grouping diseases based on their shared genetic background. To this end, an upcoming cross-disorder study of eight major psychiatric phenotypes has identified three major clusters of psychiatric disorders through genetic structural equation modeling: the first containing compulsive disorders (anorexia nervosa, obsessive-compulsive disorder and Tourette syndrome), the second consisting of mood- and psychotic disorders (depression, bipolar disorder and schizophrenia) and the third characterized by developmental disorders (autism spectrum disorder, attention-deficit hyperactivity disorder, Tourette syndrome) [75]. Besides, given the complexity of the psychiatric disorders, a classification system in which patients are diagnosed based on quantitative measures that reflect brain activity, behavior and physiology (instead of DSM-5 classification criteria) may be preferred to enhance interrater reliability and thus advance precision medicine [76,77]. As such, genetics can first of all make a contribution to this through the discovery of genetic variants influencing these quantitative variables, and second through the implementation of the genetic knowledge from these studies into predictive models (combining PRSs and relevant measures) to improve diagnostic accuracy in the individual patient and to identify homogenous subgroups of patients [68,77]. Subsequently, subgroups can be included in new genome-wide studies, further unraveling the genetic etiology of psychiatric disease and allowing for even more refinement in the diagnostic landscape. Besides this future perspective for genetics in psychiatry, GWASs based on the current clinical diagnostic system have the ability to identify genes that are shared among psychiatric diseases [71], identify pathways that are generally involved in psychiatric disorders [9], or perform drug repurposing studies to discover substances that interact with pleiotropic gene targets that can ultimately lead to the development of novel drugs that are effective in multiple psychiatric disorders [78,79].

In contrast to psychiatry, neurology is a field where a neuropathologic basis of diseases has often been defined and exams may produce more objective, reproducible and consistent findings [74]. It is striking that no genetic correlations have yet been identified between neurological disorders, except between subtypes of the same disorder (for example between focal and generalized

epilepsy and between migraine with and without aura) [73]. A plausible explanation for this is that neurological phenotypes are clearly demarcated by unique pathological hallmarks and biological processes. Nonetheless, GWAS sample sizes for neurological diseases have not yet reached those numbers as are often seen in psychiatric diseases nowadays, limiting the power of cross-disorder studies when genetic correlations are lower and more subtle [80].

It can be argued that the clear separation between psychiatry and neurology has restricted the recognition of important comorbidities and thus advancement of patient care and treatment [81]. Genetics has the ability to provide strong evidence for the sharing of underlying biology between neurological and psychiatric disease and bridge the gap between these two disciplines, as further illustrated by ***chapter 6***.

8.3.2 Genetic correlation between ALS and schizophrenia supports clinical overlap of cognitive symptoms

ALS is broadly recognized as a degenerative disease of the motor system, but includes involvement of mild cognitive impairment and executive deficits in a substantial subgroup of patients (~34 %). These cognitive symptoms are characterized by poor insight and pervasive deficits in frontal executive tests and additional change in personality, irritability and obsessions. A smaller subgroup of patients also meets criteria for frontotemporal dementia (FTD) (~14%) [82,83] and a continuum of ALS with mild cognitive impairment to progressive FTD has been proposed [84]. The rate of cognitive impairment positively correlates with disease stage in ALS and patients with final stage ALS report a higher level of psychosis compared to other disease stages, pointing at an increased neuropsychiatric burden as the disease progresses [85]. The presence of cognitive symptoms in ALS impacts both patients and their families and therefore requires further attention. Cognitive dysfunction furthermore comprises a major symptom category in psychiatric diseases including schizophrenia [86,87]. In addition, a previous cross-disorder study of ALS and FTD had already identified a modest genetic correlation with pleiotropic effects at *C9orf72* and *UNC13A* [88]. However, the genome-wide correlation between ALS and psychiatric disease had not been studied before until well-powered GWASs became available [3,89,90].

In ***chapter 6***, the genetic correlation between ALS and schizophrenia was estimated at 14.3%. This reflects a subtle but significant sharing of common variant heritability. Given the relatively low SNP-based heritability of ALS (~ 8.5%) and the fact that the genetic correlation is relative to this heritability, the estimate from our study is sensible. The genetic correlation estimate was supported by PRS analysis. The low explained variance of schizophrenia PRSs in ALS (0.12%) can be interpreted by the scores being based on a subset of SNPs from a current schizophrenia GWAS in which there is still missing heritability (i.e. the effect estimates of SNPs used for PRS calculation are not without error), that these PRSs can never explain more variance in ALS than its estimated SNP-based heritability, and that the genetic correlation is moderate [91]. Nonetheless, these findings pointed at significant sharing of genetic risk between ALS and schizophrenia and thus shared underlying biology. Employing a combined conditional false discovery rate (cFDR) analysis, we pinpointed possible pleiotropic variants involved in ALS. The discovered pleiotropy-informed risk loci from this analysis are suggestive of novel ALS loci. However, most of these are still in need of replication in an independent sample. Interestingly though, after publication of this study the cFDR locus at *TNIP1* had been discovered in a cross-ethnic ALS GWAS of European and Chinese cohorts and was replicated in independent Australian cohort [92]. Additionally, a locus highlighted by a more lenient cFDR threshold (*KIF5A*) was discovered in a meta-analysis of the ALS GWAS used in our study [89] with an additional cohort of cases [93]. This gene has furthermore been found to carry rare mutations in the region coding for the C-terminus of the protein [94], although it is currently unclear whether the common variant signal observed in GWAS tags this rare variation. Overall, discovery and replication of these suggestive loci further strengthen the robustness of findings from the cFDR analysis, of which other loci might also be discovered or replicated in future GWASs.

The estimated genetic correlation suggests that co-occurrence should be observed more often in clinical practice. However, our calculations have shown that lifetime risk for both phenotypes together is low considering the estimated correlation (1 in 34,336) and that a large (population-based) cohort of ALS patients (n = 16,488) with detailed lifetime clinical information is required to detect significant co-occurrence. These rates are also reflected in a low number of case reports of co-occurring ALS and schizophrenia [95]. In addition, the

observed genetic correlation was unlikely to be caused by misdiagnosis. Although ALS and schizophrenia have different ages of onset, we hypothesized that early onset FTD-ALS could have been misdiagnosed as schizophrenia in the respective GWAS [3]. This scenario was ruled out as no genetically heterogeneous patient subgroup carrying more ALS risk alleles than expected by chance was identified in the schizophrenia GWAS [96].

Remarkably, clinical classification systems for ALS (such as the El Escorial criteria) currently do not incorporate cognitive and behavioral phenotypes [97]. The clinical evidence [82,83,98] combined with findings from genetics [90] urges those involved in creating such classification systems to include these features into new systems to improve diagnosis of ALS patients suffering from neuropsychiatric problems and to provide better guidance for clinical follow-up and treatment of these symptoms during the course of disease [99]. Furthermore, the ascertainment of ALS patients suffering cognitive symptoms provides opportunities for future (genome-wide) studies, possibly in combined analyses with psychiatric disorders, which might lead to novel insights into genetic risk factors underlying cognitive impairment in ALS and psychiatric disease. In that context, it is also interesting to stratify on *C9orf72* status (repeat carrier or non-carrier) or take into account the repeat length of this expansion in light of the larger burden of psychiatric disease in families of *C9orf72* positive ALS patients [98]. In addition, the possible relevance of *C9orf72* for schizophrenia was highlighted in our cFDR analysis where this established gene for ALS also showed nominal significant association in schizophrenia. This is interesting in light of the earlier described role of a repeat expansion in the first intron of *C9orf72* in schizophrenia cases with a positive family history of neurodegenerative disease [100]. Additionally, a cross-disorder analysis of ALS, FTD and schizophrenia using the latest methodology would be valuable, but this possibility is still limited by the relatively low GWAS sample sizes for FTD [101].

Although the hypothesis that the shared genetic risk between ALS and schizophrenia reflects the involvement of cognitive processes is very plausible, we cannot conclusively say that variants driving the shared heritability play a role in cognition. In addition, only genetic overlap at common variants was considered. Rare variation is important in both ALS [89] and schizophrenia [102]

and future studies might look into correlations of the rare variant architectures. The absence of genetic correlation between ALS and other psychiatric diseases is counterintuitive given the high sharing of genetic risk with schizophrenia [5,72,73]. Lower sample sizes of the GWAS datasets used for those disorders might play a role and it is therefore not unlikely that future studies will identify significant genetic correlations between ALS and other psychiatric diseases.

8.3.3 Absence of shared genetic risk between ALS and epilepsy

Hyperexcitability in peripheral and central (cortical) motor neurons induces cramps and fasciculations in over half of ALS patients, often resulting in pain and thus requiring treatment [103]. Furthermore, neuronal hyperexcitability and cramps are linked to shorter survival in ALS. Overall, treatment strategies have little benefit for patients [104], although the anti-epileptic drug retigabine has recently shown acute reversal of abnormal excitability patterns in ALS patients [105]. Furthermore, a novel literature-based hypothesis has challenged the pathogenic role of hyperexcitability in ALS by postulating it as a homeostatic response to an overstimulated inhibitory nervous system [106]. This has even led to off-label treatment of a small group of patients with penicillin and hydrocortisone, which block the inhibitory nervous system [107]. Overall, these developments made a genetic cross-disorder study of ALS and epilepsy to assess their possible shared genetic and biological underpinnings a timely effort.

In contrast to shared common genetic risk in ALS and schizophrenia, this was not found between ALS and epilepsy in ***chapter 7***. Although there was no a priori evidence of common variant overlap between the two diseases, there have been case-examples where genes associated with ALS or other motor neuron disease have been specifically associated with progressive myoclonic epilepsy (PME). For example, the recently discovered *KIF5A* gene in ALS [93] was also reported for a de novo mutation in relation to severe congenital myoclonic seizures [108]. Furthermore, a rare familial repeat expansion in *C9orf72* has been linked to PME [109]. In addition to the non-significant genetic correlation, we also showed differences in the genetic architectures of ALS and epilepsy. Whereas for ALS the there is an important role for SNPs in the lower MAF spectrum (0.01-0.10), and the contribution to heritability decreases when MAF increases (shown by van Rheenen *et al.* [89] and in ***chapter 7***), there was

no predominant role for rarer SNPs in the heritability of epilepsy (including subtypes) as determined from currently available data. This does however not imply that rarer SNPs do not play a role in epilepsy, and likewise that common variants are unimportant in ALS. Future GWAS and whole-genome sequencing efforts in both diseases will uncover more of the lower-frequency variants in both diseases and allow to explore genetic similarities in this underexplored spectrum of genetic variation.

No co-occurrence of ALS and epilepsy has been highlighted by epidemiological studies. A retrospective study of medical records in over 600 ALS patients did not find a single co-occurrence of ALS and epilepsy, although this information was collected during the initial visit of the patient and thus reflected premorbid epilepsy rather than seizures developed during the course of disease [110]. For other forms of motor neuron disease like spinal muscular atrophy (SMA), comorbidity with progressive myoclonic epilepsy, where a specific single-gene mutation of *ASAH1* was causal, has been reported [111,112]. Additionally, other neurodegenerative diseases such as Alzheimer's disease or vascular dementia have a higher reported incidence of epilepsy with risk for seizures correlated to disease duration [113]. Nonetheless, this latter example might not be explained by shared common genetic risk given the absence of significant genetic correlations of epilepsy with other neurological diseases [73]. However, this study used a smaller GWAS of epilepsy [114] and did not include ALS. The study presented in ***chapter 7*** is therefore current largest and most comprehensive cross-disorder genome-wide analysis between ALS and epilepsy.

One pleiotropic locus (with corresponding directions of effect in both diseases) was found near the Bone Morphogenetic Protein 8A (*BMP8A*) and Microtubule-actin cross-linking factor 1 (*MACF1*) genes in meta-analysis of ALS and epilepsy. *MACF1* has a role in ciliogenesis [115], which is an interesting observation since pathogenic mutations in genes involved in ciliogenesis (*NEK1*) are already known in ALS [116]. The BMP gene family have a hypothesized role in ALS and motor axon navigation from animal studies [117,118]. Although intriguing, larger genome-wide studies in ALS and epilepsy are required to definitively establish whether the association at this locus is true for one or both diseases.

8.3.4 Methodological considerations in cross-disorder studies

With the public availability of summary-level results for almost all (recently) published GWASs and the development of fast and powerful tools that only require these data as input for genetic correlation estimation, such as linkage disequilibrium score regression (LDSC) [72,119] (and its online implementation in LD Hub [120]), the assessment of shared genetic risk has become standard practice in complex genetics. In this thesis, estimation of genetic correlation was based on both LDSC [72] (***chapter 6*** and ***chapter 7***) and GCTA-REML [121] (***chapter 7*** only). In contrast to LDSC, GCTA-REML uses individual-level genotypes and has higher computational requirements. Nonetheless, a study comparing both methods has revealed that accuracy of genetic correlation estimates from LDSC is sensitive to heterogeneity in input data and LD scores and recommends the use of both LDSC and GCTA-REML [122]. In the study of ALS and schizophrenia, it was not possible to use GCTA-REML due to limited access to individual-level data. Therefore, we aimed to improve the reliability of the LDSC through careful per-stratum quality control of data in the ALS GWAS [89], which allowed to provide a more conservative and accurate estimate of genetic correlation. In addition, PRSs were used to support genetic correlation estimates (applied in both ***chapter 6*** and ***chapter 7***, where in the latter the term genomic risk score, GRS, with similar definition was used) as they are expected to associate with the target phenotype in case of a substantial genetic correlation. Both cross-disorder genetic correlation and PRS analysis provide a genome-wide estimate and do therefore not pinpoint loci or genomic regions that are driving the genetic correlation. Novel methodology allows to partition the genome based on LD and estimate shared heritability for small genomic regions, although the efficiency of such an approach has thus far only been shown for polygenic quantitative traits [123]. Other approaches to identify specific shared loci include widely applied inverse variance-weighted fixed effects meta-analysis (only applicable when shared samples are excluded) [124], or derived approaches that require only summary-level data and have the ability to correct for confounding due to sample overlap [125,126].

A final and important methodological consideration for the presented cross-disorder analyses is that a pleiotropic signal at a locus does not naturally imply that two diseases share a causal variant at that locus. Apart from true biological pleiotropy, where a pleiotropic locus tags one or more causal variants in the

same gene, there are other possible scenarios [127]. In ***chapter 5*** we mentioned mediated and spurious pleiotropy. A method to distinguish mediated pleiotropy from true biological pleiotropy and to establish causal links between intermediate phenotypes and diseases is Mendelian randomization (MR) [128], which is being used more frequently in complex genetics. As causation was suspected to be unlikely between the diseases studied in ***chapter 6*** and ***chapter 7***, we did not perform MR analyses. Extensive fine mapping of established pleiotropic loci to locate causal variants can be used to distinguish biological and spurious pleiotropy [127].

Overall, in the cross-disorder studies presented in this thesis the combination of a range of approaches has resulted in cohesive evidence for the presence or absence of a shared genetic background.

8.4 Final conclusions from this thesis

Multiple psychiatric and neurological diseases have been studied in this thesis, which have in common that a complex genetic architecture composed of highly frequent variants with low effect underlies part of their liability. The aim of this thesis was to provide further insight into the genetic risk underlying these diseases, the biological mechanisms at play and the possibility of genetic similarities.

Based on the presented research, the following conclusions can now be drawn:

- Genetic variants associated with risk for schizophrenia accumulate in neuronal and synaptic pathways and more specifically in post-synaptic signaling cascades of the dopaminergic, cholinergic and glutamatergic system.

- Multivariate analysis of biochemical and behavioral phenotypes related to stress and post-trauma psychopathology discovered a locus possibly involved in the stress response and located near genes (*INTS8*, *TP53INP1*) that are strongly associated with other stress-related phenotypes.

- Polygenic risk scores capturing the combined effect of variants associated with posttraumatic stress disorder or depression did not predict the development of post-deployment psychopathology in a military cohort.

- Genetic correlation and pleiotropy are abundant in psychiatric disease and cross-disorder analyses in this field help to advance diagnostics, prognosis, and treatment.

- Amyotrophic lateral sclerosis genetically correlates with schizophrenia, which implies shared biology. This is the first and, until now, only significant genetic correlation between a neurodegenerative and psychiatric disease.

- Based on currently available datasets, there is no shared common genetic risk between amyotrophic lateral sclerosis and epilepsy.

8.5 References

1. Kane JM, Correll CU. Past and present progress in the pharmacologic treatment of schizophrenia. J Clin Psychiatry 2010;71(9):1115–24.
2. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011;43(10):969–76.
3. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511(7510):421–7.
4. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 2018;199:441.
5. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013;45(9):984–94.
6. Schijven D, Kofink D, Tragante V, Verkerke M, Pulit SL, Kahn RS, et al. Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling. Schizophr Res 2018;
7. Forero DA, Herteleer L, De Zutter S, Norrback KF, Nilsson LG, Adolfsson R, et al. A network of synaptic genes associated with schizophrenia and bipolar disorder. Schizophr Res 2016;172(1-3):68–74.
8. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA, et al. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol Psychiatry 2012;17(10):996–1006.
9. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci 2015;18(2):199–209.
10. Hertzberg L, Katsel P, Roussos P, Haroutunian V, Domany E. Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia. Schizophr Res 2015;164(1-3):92–9.
11. Duncan LE, Holmans PA, Lee PH, O'Dushlaine CT, Kirby AW, Smoller JW, et al. Pathway analyses implicate glial cells in schizophrenia. PLoS One 2014;9(2):e89441.
12. Trépanier MO, Hopperton KE, Mizrahi R, Mechawar N, Bazinet RP. Postmortem evidence of cerebral inflammation in schizophrenia: a systematic review. Mol Psychiatry 2016;21(8):1009–26.
13. Lee H, Brott BK, Kirkby LA, Adelson JD, Cheng S, Feller MB, et al. Synapse elimination and learning rules co-regulated by MHC class I H2-Db. Nature 2014;509(7499):195–200.
14. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. Nature 2016;advance online publication.
15. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27(12):1739–40.
16. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30.
17. Pers TH. Gene set analysis for interpreting genetic studies. Hum Mol Genet 2016;25(R2):R133–40.
18. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nat Rev Genet 2016;17(6):353–64.
19. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol 2015;11(4):e1004219.
20. Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D, DIAGRAM Consortium, et al. Common Inherited Variation in Mitochondrial Genes Is Not Enriched for Associations with Type 2 Diabetes or Related Glycemic Traits. PLoS Genet 2010;6(8):e1001058.
21. Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics 2012;28(13):1797–9.

22. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 2013;45(10):1150–9.
23. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet 2008;4(10):e1000214.
24. Lajoie BR, Dekker J, Kaplan N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. Methods 2015;72:65–75.
25. Won H, la Torre-Ubieta de L, Stein JL, Parikshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 2016;538(7626):523–7.
26. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 2017;550(7675):204–13.
27. GTEx Consortium, Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45(6):580–5.
28. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489(7414):57–74.
29. Zhou X, Li D, Zhang B, Lowdon RF, Rockweiler NB, Sears RL, et al. Epigenomic annotation of genetic variants using the Roadmap Epigenome Browser. Nat Biotechnol 2015;33(4):345–6.
30. Breen G, Li Q, Roth BL, O'Donnell P, Didriksen M, Dolmetsch R, et al. Translating genome-wide association findings into new therapeutics for psychiatry. Nat Neurosci 2016;19(11):1392–6.
31. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA, et al. Genetic identification of brain cell types underlying schizophrenia. Nat Genet 2018;50(6):825–33.
32. Quadrato G, Brown J, Arlotta P. The promises and challenges of human brain organoids as models of neuropsychiatric disease. Nat Med 2016;22(11):1220–8.
33. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5 ed. Washington, D.C.: 2013.
34. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, et al. DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. Am J Psychiatry 2013;170(1):59–70.
35. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PLoS One 2013;8(10):e76295.
36. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017;101(1):5–22.
37. Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nat Rev Neurosci 2006;7(10):818–27.
38. Kok BC, Herrell RK, Thomas JL, Hoge CW. Posttraumatic stress disorder associated with combat service in Iraq or Afghanistan: reconciling prevalence differences between studies. J Nerv Ment Dis 2012;200(5):444–50.
39. Koenen KC, Ratanatharathorn A, Ng L, McLaughlin KA, Bromet EJ, Stein DJ, et al. Posttraumatic stress disorder in the World Mental Health Surveys. Psychol Med 2017;47(13):2260–74.
40. Lapierre CB, Schwegler AF, Labauve BJ. Posttraumatic stress and depression symptoms in soldiers returning from combat operations in Iraq and Afghanistan. J Trauma Stress 2007;20(6):933–43.
41. Logue MW, Baldwin C, Guffanti G, Melista E, Wolf EJ, Reardon AF, et al. A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. Mol Psychiatry 2013;18(8):937–42.
42. Nievergelt CM, Maihofer AX, Mustapic M, Yurgil KA, Schork NJ, Miller MW, et al. Genomic predictors of combat stress vulnerability and resilience in U.S. Marines: A genome-wide association study across multiple ancestries implicates PRTFDC1 as a potential PTSD gene. Psychoneuroendocrinology 2015;51:459–71.

43. Stein MB, Chen CY, Ursano RJ, Cai T, Gelernter J, Heeringa SG, et al. Genome-wide Association Studies of Posttraumatic Stress Disorder in 2 Cohorts of US Army Soldiers. JAMA Psychiatry 2016;
44. Duncan LE, Ratanatharathorn A, Aiello AE, Almli LM, Amstadter AB, Ashley-Koch AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. Mol Psychiatry 2018;23(3):666–73.
45. Xie P, Kranzler HR, Yang C, Zhao H, Farrer LA, Gelernter J. Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. Biol Psychiatry 2013;74(9):656–63.
46. Almli LM, Stevens JS, Smith AK, Kilaru V, Meng Q, Flory J, et al. A genome-wide identified risk variant for PTSD is a methylation quantitative trait locus and confers decreased cortical activation to fearful faces. Am J Med Genet B Neuropsychiatr Genet 2015;168B(5):327–36.
47. Nievergelt CM, Ashley-Koch AE, Dalvie S, Hauser MA, Morey RA, Smith AK, et al. Genomic Approaches to Posttraumatic Stress Disorder: The Psychiatric Genomic Consortium Initiative. Biol Psychiatry 2018;83(10):831–9.
48. Kraft P, Zeggini E, Ioannidis JPA. Replication in Genome-Wide Association Studies. Stat Sci 2009;24(4):561–73.
49. Gray GC, Chesbrough KB, Ryan MAK, Amoroso P, Boyko EJ, Gackstetter GD, et al. The millennium Cohort Study: a 21-year prospective cohort study of 140,000 military personnel. Mil Med 2002;167(6):483–8.
50. Baker DG, Nash WP, Litz BT, Geyer MA, Risbrough VB, Nievergelt CM, et al. Predictors of risk and resilience for posttraumatic stress disorder among ground combat Marines: methods of the Marine Resiliency Study. Prev Chronic Dis 2012;9:E97.
51. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. Nat Rev Genet 2016;17(3):129–45.
52. Escott-Price V, Bellenguez C, Wang L-S, Choi S-H, Harold D, Jones L, et al. Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. PLoS One 2014;9(6):e94661.
53. Miller MW, Lin AP, Wolf EJ, Miller DR. Oxidative Stress, Inflammation, and Neuroprogression in Chronic PTSD. Harv Rev Psychiatry 2018;26(2):57–69.
54. Yaffe K, Vittinghoff E, Lindquist K, Barnes D, Covinsky KE, Neylan T, et al. Posttraumatic stress disorder and risk of dementia among US veterans. Archives of general psychiatry 2010;67(6):608–13.
55. Mohamed AZ, Cumming P, Srour H, Gunasena T, Uchida A, Haller CN, et al. Amyloid pathology fingerprint differentiates post-traumatic stress disorder and traumatic brain injury. Neuroimage Clin 2018;19:716–26.
56. Porter HF, O'Reilly PF. Multivariate simulation framework reveals performance of multi-trait GWAS methods. Sci Rep 2017;7(1):38837.
57. Magi R, Suleimanov YV, Clarke GM, Kaakinen M, Fischer K, Prokopenko I, et al. SCOPA and META-SCOPA: software for the analysis and aggregation of genome-wide association studies of multiple correlated phenotypes. BMC Bioinformatics 2017;18(1):25.
58. O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin M-R, et al. MultiPhen: Joint Model of Multiple Phenotypes Can Increase Discovery in GWAS. PLoS One 2012;7(5):e34861.
59. Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, et al. Meta-analysis of genome-wide association studies of anxiety disorders. Mol Psychiatry 2016;21(10):1391–9.
60. Olff M, Langeland W, Draijer N, Gersons BPR. Gender differences in posttraumatic stress disorder. Psychol Bull 2007;133(2):183–204.
61. Gonda X, Hullam G, Antal P, Eszlari N, Petschner P, Hökfelt TG, et al. Significance of risk polymorphisms for depression depends on stress exposure. Sci Rep 2018;8(1):3946.
62. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 2007;17(10):1520–8.
63. Dudbridge F. Power and Predictive Accuracy of Polygenic Risk Scores. PLoS Genet 2013;9(3):e1003348.

64. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018;50(5):668–81.
65. Eekhout I, Reijnen A, Vermetten E, Geuze E. Post-traumatic stress symptoms 5 years after military deployment to Afghanistan: an observational cohort study. Lancet Psychiatry 2016;3(1):58–64.
66. Rice F, Riglin L, Thapar AK, Heron J, Anney R, O'Donovan MC, et al. Characterizing Developmental Trajectories and the Role of Neuropsychiatric Genetic Risk Variants in Early-Onset Depression. JAMA Psychiatry 2018;
67. Peyrot WJ, Milaneschi Y, Abdellaoui A, Sullivan PF, Hottenga JJ, Boomsma DI, et al. Effect of polygenic risk scores on depression in childhood trauma. Br J Psychiatry 2014;205(2):113–9.
68. Bogdan R, Baranger DAA, Agrawal A. Polygenic Risk Scores in Clinical Psychology: Bridging Genomic Risk to Individual Differences. Annu Rev Clin Psychol 2018;14(1):119–57.
69. Meyers JL, Cerdá M, Galea S, Keyes KM, Aiello AE, Uddin M, et al. Interaction between polygenic risk for cigarette use and environmental exposures in the Detroit Neighborhood Health Study. Transl Psychiatry 2013;3(8):e290–0.
70. Visscher PM, Yang J. A plethora of pleiotropy across complex traits. Nat Genet 2016;48(7):707–8.
71. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013;381(9875):1371–9.
72. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47(11):1236–41.
73. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science 2018;360(6395):eaap8757.
74. Price BH, Adams RD, Coyle JT. Neurology and psychiatry: closing the great divide. Neurology 2000;54(1):8–14.
75. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. bioRxiv 2019;
76. O'Donovan MC, Owen MJ. The implications of the shared genetics of psychiatric disorders. Nat Med 2016;22(11):1214–9.
77. Insel TR, Cuthbert BN. Medicine. Brain disorders? Precisely. Science 2015;348(6234):499–500.
78. Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. Sci Rep 2017;7(1):12460.
79. Lam M, Trampush JW, Yu J, Knowles E, Davies G, Liewald DC, et al. Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. Cell Rep 2017;21(9):2597–613.
80. Visscher PM, Hemani G, Vinkhuyzen AAE, Chen G-B, Lee SH, Wray NR, et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet 2014;10(4):e1004269.
81. Fitzgerald M. Do psychiatry and neurology need a close partnership or a merger? BJPsych Bull 2015;39(3):105–7.
82. Phukan J, Pender NP, Hardiman O. Cognitive impairment in amyotrophic lateral sclerosis. Lancet Neurol 2007;6(11):994–1003.
83. Phukan J, Elamin M, Bede P, Jordan N, Gallagher L, Byrne S, et al. The syndrome of cognitive impairment in amyotrophic lateral sclerosis: a population-based study. J Neurol Neurosurg Psychiatry 2012;83(1):102–8.
84. Strong MJ, Abrahams S, Goldstein LH, Woolley S, Mclaughlin P, Snowden J, et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. Amyotroph Lateral Scler Frontotemporal Degener 2017;18(3-4):153–74.

85. Crockford C, Newton J, Lonergan K, Chiwera T, Booth T, Chandran S, et al. ALS-specific cognitive and behavior changes associated with advancing disease stage in ALS. Neurology 2018;91(15):e1370–80.
86. Kahn RS, Sommer IE, Murray RM, Meyer-Lindenberg A, Weinberger DR, Cannon TD, et al. Schizophrenia. Nat Rev Dis Primers 2015;:15067.
87. Kahn RS, Keefe RS. Schizophrenia is a cognitive illness: time for a change in focus. JAMA Psychiatry 2013;70(10):1107–12.
88. Diekstra FP, Van Deerlin VM, van Swieten JC, Al-Chalabi A, Ludolph AC, Weishaupt JH, et al. C9orf72 and UNC13A are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: A genome-wide meta-analysis. Ann Neurol 2014;76(1):120–33.
89. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat Genet 2016;48(9):1043–8.
90. McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun 2017;8:14774.
91. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet 2013;14(7):507–15.
92. Benyamin B, He J, Zhao Q, Gratten J, Garton F, Leo PJ, et al. Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. Nat Commun 2017;8(1):611.
93. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron 2018;97(6):1268–1283.e6.
94. Brenner D, Yilmaz R, Müller K, Grehl T, Petri S, Meyer T, et al. Hot-spot KIF5A mutations cause familial ALS. Brain 2018;141(3):688–97.
95. Howland RH. Schizophrenia and amyotrophic lateral sclerosis. Compr Psychiatry 1990;31(4):327–36.
96. Han B, Pouget JG, Slowikowski K, Stahl E, Lee CH, Diogo D, et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nat Genet 2016;48(7):803–10.
97. Hardiman O, Al-Chalabi A, Chiò A, Corr EM, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. Nat Rev Dis Primers 2017;3:17071.
98. Byrne S, Heverin M, Elamin M, Bede P, Lynch C, Kenna K, et al. Aggregation of neurologic and neuropsychiatric disease in amyotrophic lateral sclerosis kindreds: A population-based case–control cohort study of familial and sporadic amyotrophic lateral sclerosis. Ann Neurol 2013;74(5):699–708.
99. Al-Chalabi A, Hardiman O, Kiernan MC, Chiò A, Rix-Brooks B, van den Berg LH. Amyotrophic lateral sclerosis: moving towards a new classification system. Lancet Neurol 2016;15(11):1182–94.
100. Galimberti D, Reif A, Dell'Osso B, Kittel-Schneider S, Leonhard C, Herr A, et al. The C9ORF72 hexanucleotide repeat expansion is a rare cause of schizophrenia. Neurobiol Aging 2014;35(5):1214.e7–1214.e10.
101. Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JBJ, et al. Frontotemporal dementia and its subtypes: a genome-wide association study. Lancet Neurol 2014;13(7):686–99.
102. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506(7487):185–90.
103. Bae JS, Simon NG, Menon P, Vucic S, Kiernan MC. The puzzling case of hyperexcitability in amyotrophic lateral sclerosis. J Clin Neurol 2013;9(2):65–74.
104. Baldinger R, Katzberg HD, Weber M. Treatment for cramps in amyotrophic lateral sclerosis/motor neuron disease. Cochrane Database Syst Rev 2012;330(4):CD004157.

105. Kovalchuk MO, Heuberger JAAC, Sleutjes BTHM, Ziagkos D, van den Berg LH, Ferguson TA, et al. Acute Effects of Riluzole and Retigabine on Axonal Excitability in Patients With Amyotrophic Lateral Sclerosis: A Randomized, Double-Blind, Placebo-Controlled, Crossover Trial. Clin Pharmacol Ther 2018;9(Pt 3):65.
106. Tuk B. Overstimulation of the inhibitory nervous system plays a role in the pathogenesis of neuromuscular and neurological diseases: a novel hypothesis. F1000Res 2016;5(1435).
107. Tuk B, Jousma H, Gaillard PJ. Treatment with penicillin G and hydrocortisone reduces ALS-associated symptoms: a case series of three patients. F1000Res 2017;6:410.
108. Rydzanicz M, Jagła M, Kosinska J, Tomasik T, Sobczak A, Pollak A, et al. KIF5A de novo mutation associated with myoclonic seizures and neonatal onset progressive leukoencephalopathy. Clin Genet 2017;91(5):769–73.
109. van den Ameele J, Jedlickova I, Pristoupilova A, Sieben A, Van Mossevelde S, Ceuterick-de Groote C, et al. Teenage-onset progressive myoclonic epilepsy due to a familial C9orf72 repeat expansion. Neurology 2018;90(8):e658–63.
110. Tartaglia MC, Rowe A, Findlater K, Orange JB, Grace G, Strong MJ. Differentiation Between Primary Lateral Sclerosis and Amyotrophic Lateral Sclerosis. Arch Neurol 2007;64(2):232–6.
111. Rubboli G, Veggiotti P, Pini A, Berardinelli A, Cantalupo G, Bertini E, et al. Spinal muscular atrophy associated with progressive myoclonic epilepsy: A rare condition caused by mutations in ASAH1. Epilepsia 2015;56(5):692–8.
112. Zhou J, Tawk M, Tiziano FD, Veillet J, Bayes M, Nolent F, et al. Spinal muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in ASAH1. Am J Hum Genet 2012;91(1):5–14.
113. Imfeld P, Bodmer M, Schuerch M, Jick SS, Meier CR. Seizures in patients with Alzheimer's disease or vascular dementia: a population-based nested case-control analysis. Epilepsia 2013;54(4):700–7.
114. The International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol 2014;13(9):893–903.
115. May-Simera HL, Gumerson JD, Gao C, Campos M, Cologna SM, Beyer T, et al. Loss of MACF1 Abolishes Ciliogenesis and Disrupts Apicobasal Polarity Establishment in the Retina. Cell Rep 2016;17(5):1399–413.
116. Kenna KP, van Doormaal PTC, Dekker AM, Ticozzi N, Kenna BJ, Diekstra FP, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat Genet 2016;
117. Shijo T, Warita H, Suzuki N, Ikeda K, Mitsuzawa S, Akiyama T, et al. Antagonizing bone morphogenetic protein 4 attenuates disease progression in a rat model of amyotrophic lateral sclerosis. Exp Neurol 2018;307:164–79.
118. Jardin N, Giudicelli F, Martín Ten D, Vitrac A, De Gois S, Allison R, et al. BMP- and neuropilin 1-mediated motor axon navigation relies on spastin alternative translation. Development 2018;145(17):dev162701.
119. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Psychiatric GWAS Consortium Bipolar Disorder Working Group, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47(3):291–5.
120. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock, P. C., et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 2017;33(2):272–9.
121. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics 2012;28(19):2540–2.
122. Ni G, Moser G, Psychiatric GWAS Consortium Bipolar Disorder Working Group, Wray NR, Lee SH. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. Am J Hum Genet 2018;102(6):1185–94.
123. Shi H, Mancuso N, Spendlove S, Pasaniuc B. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. Am J Hum Genet 2017;101(5):737–51.

124. de Bakker PIW, Neale BM, Daly MJ. Meta-analysis of genome-wide association studies. Cold Spring Harb Protoc 2010;2010(6):pdb.top81–1.
125. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet 2018;9:283.
126. Liley J, Wallace C. A Pleiotropy-Informed Bayesian False Discovery Rate Adapted to a Shared Control Design Finds New Disease Associations From GWAS Summary Statistics. PLoS Genet 2015;11(2):e1004926.
127. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 2013;14(7):483–95.
128. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr 2016;103(4):965–78.

Appendix I

Consortium members

Project MinE ALS GWAS Consortium members

Wouter van Rheenen, Aleksey Shatunov, Russell L McLaughlin, Annelot M Dekker, Frank P Diekstra, Sara L Pulit, Rick AA van der Spek, Perry TC van Doormaal, William Sproviero, Ashley R Jones, Garth A Nicholson, Dominic B Rowe, Roger Pamphlett, Matthew C Kiernan, Denis Bauer, Tim Kahlke, Kelly Williams, Filip Eftimov, Isabella Fogh,, Nicola Ticozzi, Kuang Lin, Stéphanie Millecamps, François Salachas, Vincent Meininger, Mamede de Carvalho, Susana Pinto, Jesus S Mora, Ricardo Rojas-García, Meraida Polak, Siddharthan Chandran, Shuna Colville, Robert Swingler, Karen E Morrison, Pamela J Shaw, John Hardy, Richard W Orrell, Alan Pittman, Katie Sidle, Pietro Fratta, Andrea Malaspina, Susanne Petri, Susanna Abdulla, Carsten Drepper, Michael Sendtner, Thomas Meyer, Roel A Ophoff, Martina Wiedau-Pazos, Catherine Lomen-Hoerth, Vivianna M Van Deerlin, John Q Trojanowski, Lauren Elman, Leo McCluskey, Nazli Basak, Thomas Meitinger, Peter Lichtner, Milena Blagojevic-Radivojkov, Christian R Andres, Cindy Maurel, Gilbert Bensimon, Bernhard Landwehrmeyer, Alexis Brice, Christine AM Payan, Safa Saker-Delye, Alexandra Dürr, Nicholas Wood, Lukas Tittmann, Wolfgang Lieb, Andre Franke, Marcella Rietschel, Sven Cichon, Markus M Nöthen, Philippe Amouyel, Christophe Tzourio, Jean-François Dartigues, Andre G Uitterlinden, Fernando Rivadeneira, Karol Estrada, Albert Hofman, Charles Curtis, Anneke J van der Kooi, Marianne de Visser, An Goris, Markus Weber, Christopher E Shaw, Bradley N Smith, Orietta Pansarasa, Cristina Cereda, Roberto Del Bo, Giacomo P Comi, Sandra D'Alfonso, Cinzia Bertolin, Gianni Sorarù, Letizia Mazzini, Viviana Pensato, Cinzia Gellera, Cinzia Tiloca, Antonia Ratti, Andrea Calvo, Cristina Moglia, Maura Brunetti, Simon Arcuti, Rosa Capozzo, Chiara Zecca, Christian Lunetta, Silvana Penco, Nilo Riva, Alessandro Padovani, Massimiliano Filosto, Ian Blair, P Nigel Leigh, Federico Casale, Adriano Chio, Ettore Beghi, Elisabetta Pupillo, Rosanna Tortelli, Giancarlo Logroscino, John Powell, Albert C Ludolph, Jochen H Weishaupt, Wim Robberecht, Philip Van Damme, Robert H Brown, Jonathan Glass, John E Landers, Peter M Andersen, Philippe Corcia, Patrick Vourc'h, Vincenzo Silani, Michael A van Es, R Jeroen Pasterkamp, Cathryn M Lewis, Gerome Breen, Ammar Al-Chalabi, Leonard H van den Berg, Jan H Veldink, PARALS registry, SLALOM group, SLAP registry, FALS Sequencing Consortium, SLAGEN Consortium, NNIPPS Study Group.

PARALS registry: Stefania Cammarosano, Antonio Canosa, Dario Cocito, Leonardo Lopiano, Luca Durelli, Bruno Ferrero, Antonio Bertolotto, Alessandro Mauro, Luca Pradotto, Roberto Cantello, Enrica Bersano, Dario Giobbe, Maurizio Gionco, Daniela Leotta, Lucia Appendino, Roberto Cavallo, Enrico Odddenino, Claudio Geda, Fabio Poglio, Paola Santimaria, Umberto Massazza, Angelo Villani, Roberto Conti, Fabrizio Pisano, Mario Palermo, Franco Vergnano, Paolo Provera, Maria Teresa Penza, Marco Aguggia, Nicoletta Di Vito, Piero Meineri, Ilaria Pastore, Paolo Ghiglione, Danilo Seliak, Nicola Launaro, Giovanni Astegiano, Bottacchi Edo.

SLALOM group: Francesca Gerardi, Fabrizio Rinaldi, Maria Sofia Cotelli, Luca Chiveri, Maria Cristina Guaita, Patrizia Perrone.

SLAP registry: Isabella Laura Simon, Stefano Zoccolella, Michele Zarrelli, Franco Apollo.

FALS Sequencing Consortium: Bradley N Smith, Nicola Ticozzi, Claudia Fallini, Athina Soragia Gkazi, Simon Topp, Jason Kost, Emma L Scotter, Kevin P Kenna, Pamela Keagle, Jack W Miller, Cinzia Tiloca, Caroline Vance, Eric W Danielson, Claire Troakes, Claudia Colombrita, Safa Al-Sarraj, Elizabeth A Lewis, Andrew King, Daniela Calini, Viviana Pensato, Barbara Castellotti, Jacqueline de Belleroche, Frank Baas, Anneloor LMA ten Asbroek, Peter C Sapp, Diane McKenna-Yasek, Russell L McLaughlin, Meraida Polak, Seneshaw Asress, Jesús Esteban-Pérez, José Luis Muñoz-Blanco, Sandra D'Alfonso, Letizia Mazzini, Giacomo P Comi, Roberto Del Bo, Mauro Ceroni, Stella Gagliardi, Giorgia Querin, Cinzia Bertolin, Wouter van Rheenen, Frank P Diekstra, Rosa Rademakers, Marka van Blitterswijk, Kevin B Boylan, Giuseppe Lauria, Stefano Duga, Stefania Corti, Cristina Cereda, Lucia Corrado, Gianni Sorarù, Kelly L Williams, Garth A Nicholson,

Ian P Blair, Claire Leblond-Manry, Guy A Rouleau, Orla Hardiman, Karen E Morrison, Jan H Veldink, Leonard H van den Berg, Ammar Al-Chalabi, Hardev Pall, Pamela J Shaw, Martin R Turner, Kevin Talbot, Franco Taroni, Alberto García-Redondo, Zheyang Wu, Jonathan D Glass, Cinzia Gellera, Antonia Ratti, Robert H Brown Jr, Vincenzo Silani, Christopher E Shaw, John E Landers.

SLAGEN Consortium: Daniela Calini, Isabella Fogh, Antonia Ratti, Vincenzo Silani, Nicola Ticozzi, Cinzia Tiloca, Barbara Castellotti, Cinzia Gellera, Viviana Pensato, Franco Taroni, Cristina Cereda, Mauro Ceroni, Stella Gagliardi, Giacomo Comi, Stefania Corti, Roberto Del Bo, Lucia Corrado, Sandra D'Alfonso, Letizia Mazzini, Elena Pegoraro, Giorgia Querin, Gianni Sorarù.

NNIPPS Study Group: William Camu, Jean Sebastien Hulot, Francois Viallet, Philippe Couratier, David Maltete, Christine Tranchant, Marie Vidailhet.

Schizophrenia Working Group of the Psychiatric Genomics Consortium members

Stephan Ripke, Benjamin M Neale, Aiden Corvin, James TR Walters, Kai-How Farh, Peter A Holmans, Phil Lee, Brendan Bulik-Sullivan, David A Collier, Hailiang Huang, Tune H Pers, Ingrid Agartz, Esben Agerbo, Margot Albus, Madeline Alexander, Farooq Amin, Silviu A Bacanu, Martin Begemann, Richard A Belliveau Jr, Judit Bene, Sarah E Bergen, Elizabeth Bevilacqua, Tim B Bigdeli, Donald W Black, Richard Bruggeman, Nancy G Buccola, Randy L Buckner, William Byerley, Wiepke Cahn, Guiqing Cai, Dominique Campion, Rita M Cantor, Vaughan J Carr, Noa Carrera, Stanley V Catts, Kimberley D Chambert, Raymond CK Chan, Ronald YL Chan, Eric YH Chen, Wei Cheng, Eric FC Cheung, Siow Ann Chong, C Robert Cloninger, David Cohen, Nadine Cohen, Paul Cormican, Nick Craddock, James J Crowley, David Curtis, Michael Davidson, Kenneth L Davis, Franziska Degenhardt, Jurgen Del Favero, Ditte Demontis, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Gary Donohoe, Elodie Drapeau, Jubao Duan, Frank Dudbridge, Naser Durmishi, Peter Eichhammer, Johan Eriksson, Valentina Escott-Price, Laurent Essioux, Ayman H Fanous, Martilias S Farrell, Josef Frank, Lude Franke, Robert Freedman, Nelson B Freimer, Marion Friedl, Joseph I Friedman, Menachem Fromer, Giulio Genovese, Lyudmila Georgieva, Ina Giegling, Paola Giusti-Rodríguez, Stephanie Godard, Jacqueline I Goldstein, Vera Golimbet, Srihari Gopal, Jacob Gratten, Lieuwe de Haan, Christian Hammer, Marian L Hamshere, Mark Hansen, Thomas Hansen, Vahram Haroutunian, Annette M Hartmann, Frans A Henskens, Stefan Herms, Joel N Hirschhorn, Per Hoffmann, Andrea Hofman, Mads V Hollegaard, David M Hougaard, Masashi Ikeda, Inge Joa, Antonio Julià, Luba Kalaydjieva, Sena Karachanak-Yankova, Juha Karjalainen, David Kavanagh, Matthew C Keller, James L Kennedy, Andrey Khrunin, Yunjung Kim, Janis Klovins, James A Knowles, Bettina Konte, Vaidutis Kucinskas, Zita Ausrele Kucinskiene, Hana Kuzelova-Ptackova, Anna K Kähler, Claudine Laurent, Jimmy Lee, S Hong Lee, Sophie E Legge, Bernard Lerer, Miaoxin Li, Tao Li, Kung-Yee Liang, Jeffrey Lieberman, Svetlana Limborska, Carmel M Loughland, Jan Lubinski, Jouko Lönnqvist, Milan Macek, Patrik KE Magnusson, Brion S Maher, Wolfgang Maier, Jacques Mallet, Sara Marsal, Manuel Mattheisen, Morten Mattingsdal, Robert W McCarley, Colm McDonald, Andrew M McIntosh, Sandra Meier, Carin J Meijer, Bela Melegh, Ingrid Melle, Raquelle I Mesholam-Gately, Andres Metspalu, Patricia T Michie, Lili Milani, Vihra Milanova, Younes Mokrab, Derek W Morris, Ole Mors, Kieran C Murphy, Robin M Murray, Inez Myin-Germeys, Bertram Müller-Myhsok, Mari Nelis, Igor Nenadic, Deborah A Nertney, Gerald Nestadt, Kristin K Nicodemus, Liene Nikitina-Zake, Laura Nisenbaum, Annelie Nordin, Eadbhard O'Callaghan, Colm O'Dushlaine, F Anthony O'Neill, Sang-Yun Oh, Ann Olincy, Line Olsen, Jim Van Os, Christos Pantelis, George N Papadimitriou, Sergi Papiol, Elena Parkhomenko, Michele T Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Diana O Perkins, Olli Pietiläinen, Jonathan Pimm, Andrew J Pocklington, Alkes Price, Ann E Pulver, Shaun M Purcell, Digby Quested, Henrik B Rasmussen, Abraham Reichenberg, Mark A Reimers, Alexander L Richards, Joshua L Roffman, Panos Roussos, Douglas M Ruderfer, Veikko Salomaa, Alan R Sanders, Ulrich Schall, Christian R Schubert, Thomas G Schulze,

Sibylle G Schwab, Edward M Scolnick, Rodney J Scott, Larry J Seidman, Jianxin Shi, Engilbert Sigurdsson, Teimuraz Silagadze, Jeremy M Silverman, Kang Sim, Petr Slominsky, Jordan W Smoller, Hon-Cheong So, Chris C A Spencer, Eli A Stahl, Hreinn Stefansson, Stacy Steinberg, Elisabeth Stogmann, Richard E Straub, Eric Strengman, Jana Strohmaier, T Scott Stroup, Mythily Subramaniam, Jaana Suvisaari, Dragan M Svrakic, Jin P Szatkiewicz, Erik Söderman, Srinivas Thirumalai, Draga Toncheva, Sarah Tosato, Juha Veijola, John Waddington, Dermot Walsh, Dai Wang, Qiang Wang, Bradley T Webb, Mark Weiser, Dieter B Wildenauer, Nigel M Williams, Stephanie Williams, Stephanie H Witt, Aaron R Wolen, Emily HM Wong, Brandon K Wormley, Hualin Simon Xi, Clement C Zai, Xuebin Zheng, Fritz Zimprich, Naomi R Wray, Kari Stefansson, Peter M Visscher, Rolf Adolfsson, Ole A Andreassen, Douglas HR Blackwood, Elvira Bramon, Joseph D Buxbaum, Anders D Børglum, Ariel Darvasi, Enrico Domenici, Hannelore Ehrenreich, Tõnu Esko, Pablo V Gejman, Michael Gill, Hugh Gurling, Christina M Hultman, Nakao Iwata, Assen V Jablensky, Erik G Jönsson, Kenneth S Kendler, George Kirov, Jo Knight, Todd Lencz, Douglas F Levinson, Qingqin S Li, Jianjun Liu, Anil K Malhotra, Steven A McCarroll, Andrew McQuillin, Jennifer L Moran, Preben B Mortensen, Bryan J Mowry, Michael J Owen, Aarno Palotie, Carlos N Pato, Tracey L Petryshen, Danielle Posthuma, Brien P Riley, Dan Rujescu, Pak C Sham, Pamela Sklar, David St Clair, Daniel R Weinberger, Jens R Wendland, Thomas Werge, Mark J Daly, Patrick F Sullivan, Michael C O'Donovan.

The International League Against Epilepsy Consortium on Complex Epilepsies members

Bassel Abou-Khalil, Pauls Auce, Andreja Avbersek, Melanie Bahlo, David J Balding, Thomas Bast, Larry Baum, Albert J Becker, Felicitas Becker, Bianca Berghuis, Samuel F Berkovic, Katja E Boysen, Jonathan P Bradfield, Lawrence C Brody, Russell J Buono, Ellen Campbell, Gregory D Cascino, Claudia B Catarino, Gianpiero L Cavalleri, Stacey S Cherny, Krishna Chinthapalli, Alison J Coffey, Alastair Compston, Antonietta Coppola, Patrick Cossette, John J Craig, Gerrit-Jan de Haan, Peter De Jonghe, Carolien G F de Kovel, Norman Delanty, Chantal Depondt, Orrin Devinsky, Dennis J Dlugos, Colin P Doherty, Christian E Elger, Johan G Eriksson, Thomas N Ferraro, Martha Feucht, Ben Francis, Andre Franke, Jacqueline A French, Saskia Freytag, Verena Gaus, Eric B Geller, Christian Gieger, Tracy Glauser, Simon Glynn, David B Goldstein, Hongsheng Gui, Youling Guo, Kevin F Haas, Hakon Hakonarson, Kerstin Hallmann, Sheryl Haut, Erin L Heinzen, Ingo Helbig, Christian Hengsbach, Helle Hjalgrim, Michele Iacomino, Andrés Ingason, Jennifer Jamnadas-Khoda, Michael R Johnson, Reetta Kälviäinen, Anne-Mari Kantanen, Dalia Kasperavičiūte, Dorothee Kasteleijn-Nolst Trenite, Heidi E Kirsch, Robert C Knowlton, Bobby P C Koeleman, Roland Krause, Martin Krenn, Wolfram S Kunz, Ruben Kuzniecky, Patrick Kwan, Dennis Lal, Yu-Lung Lau, Anna-Elina Lehesjoki, Holger Lerche, Costin Leu, Wolfgang Lieb, Dick Lindhout, Warren D Lo, Iscia Lopes-Cendes, Daniel H Lowenstein, Alberto Malovini, Anthony G Marson, Thomas Mayer, Mark McCormack, James L Mills, Nasir Mirza, Martina Moerzinger, Rikke S Møller, Anne M Molloy, Hiltrud Muhle, Mark Newton, Ping-Wing Ng, Markus M Nöthen, Peter Nürnberg, Terence J O'Brien, Karen L Oliver, Aarno Palotie, Faith Pangilinan, Sarah Peter, Slavé Petrovski, Annapurna Poduri, Michael Privitera, Rodney Radtke, Sarah Rau, Philipp S Reif, Eva M Reinthaler, Felix Rosenow, Josemir W Sander, Thomas Sander, Theresa Scattergood, Steven C Schachter, Christoph J Schankin, Ingrid E Scheffer, Bettina Schmitz, Susanne Schoch, Pak C Sham, Jerry J Shih, Graeme J Sills, Sanjay M Sisodiya, Lisa Slattery, Alexander Smith, David F Smith, Michael C Smith, Philip E Smith, Anja C M Sonsma, Doug Speed, Michael R Sperling, Bernhard J Steinhoff, Ulrich Stephani, Remi Stevelink, Konstantin Strauch, Pasquale Striano, Hans Stroink, Rainer Surges, K Meng Tan, Liu Lin Thio, G Neil Thomas, Marian Todaro, Rossana Tozzi, Maria S Vari, Eileen PG Vining, Frank Visscher, Sarah von Spiczak, Nicole M Walley, Yvonne G Weber, Zhi Wei, Judith Weisenberg, Christopher D Whelan, Peter Widdess-Walsh, Markus Wolff, Stefan Wolking, Wanling Yang, Federico Zara, Fritz Zimprich.

Appendix II

Samenvatting in het Nederlands

Genoomwijde studies op het grensvlak van diagnostische classificatie in de psychiatrie en neurologie

Omstreeks het midden van de negentiende eeuw observeerde en rapporteerde Gregor Mendel, een Tsjechische monnik, de overerving van waarneembare eigenschappen (het **fenotype***) in opeenvolgende generaties van erwtenplanten. Zijn waarnemingen stonden aan de basis van de zoektocht naar de drager van deze erfelijke informatie. Dat resulteerde ongeveer een eeuw later in de ontdekking van het molecuul **desoxyribonucleïnezuur (DNA)**. DNA bevat de genetische code, opgebouwd uit een opeenvolgende sequentie van bouwstenen: **nucleotiden** met daaraan een van de basen adenine (vaak afgekort met de letter A), thymine (T), cytosine (C) of guanine (G). Het humane **genoom** bestaat uit ongeveer drie miljard van deze bouwstenen. Hoewel deze genetische code grotendeels overeenkomt tussen mensen, bevat deze ook een aanzienlijke hoeveelheid variatie. Genetische variatie resulteert in variatie in het fenotype, omdat varianten invloed hebben op **genen** die coderen voor **eiwitten** waarmee alle biologische processen in het lichaam gereguleerd worden. Omdat het DNA, en daarmee genetische variatie, wordt doorgegeven aan nakomelingen zijn veel fenotypen **erfelijk**.

Veel van de genetische variatie resulteert in fenotypen die over het algemeen niet nadelig zijn (zoals oogkleur of lengte), maar er zijn ook varianten die het risico op ziekten vergroten. Sommige varianten hebben een dusdanig groot effect op een gen, eiwit of biologisch proces dat het dragen van één dergelijke variant al ziekte veroorzaakt. Echter, veel complexe ziekten (waaronder ook veel psychiatrische en neurologische aandoeningen) worden veroorzaakt door meerdere varianten die elk kleine effecten sorteren. Van dit soort varianten komt het **enkel-nucleotide polymorfisme** (in het Engels "single nucleotide polymorphism" en daarom vaak afgekort als SNP, uitspraak: *snip*) het meest voor in het DNA: iedere persoon draagt miljoenen SNPs en deze komen dus ook zeer frequent voor in een populatie. SNPs zijn de genetische varianten die aan de basis staan van al het beschreven onderzoek in dit proefschrift.

De genetische code is geheel in kaart gebracht, waardoor locaties van SNPs (grotendeels) bekend zijn. Relatief nieuwe technieken maken het mogelijk om de genetische code op deze locaties af te lezen. Met behulp van **genoomwijde associatiestudies (GWAS)** kan getest worden of bepaalde SNPs vaker voorkomen

*** Gemarkeerde termen** worden in de verklarende woordenlijst aan het einde van deze samenvatting toegelicht.

in combinatie met een fenotype of ziekte, door de genetische code op de afgelezen locaties tussen duizenden gezonde en zieke individuen te vergelijken. GWAS markeren kleine gebieden op het DNA (loci, enkelvoud: **locus**) die informatie kunnen bevatten over betrokken genen, met als uiteindelijk doel om biologische processen te identificeren die verstoord zijn in ziekten en daarmee het ontstaan ervan te verklaren en de ontwikkeling van medicijnen te bevorderen. Een uitgebreide introductie op de onderzoeken beschreven in dit proefschrift is te lezen in ***hoofdstuk 1***. In het grijze kader in deze samenvatting worden bovendien de in dit proefschrift bestudeerde psychiatrische en neurologische ziekten kort samengevat.

Voor schizofrenie zijn GWAS tot op heden zeer succesvol geweest en er zijn inmiddels rond de 150 loci ontdekt die significant geassocieerd zijn met de ziekte. Verrijkingsanalyses, waarbij de accumulatie van geassocieerde SNPs in sets van genen betrokken bij biologische processen wordt getest, kunnen bestaande theorieën over de biologische oorzaak van schizofrenie vanuit een genetisch perspectief ondersteunen en de generatie van nieuwe hypotheses bevorderen. In ***hoofdstuk 2*** beschrijf ik een dergelijke verrijkingsanalyse gebaseerd op de resultaten van de grootste GWASs in schizofrenie. We vonden een significante verrijking van met schizofrenie geassocieerde varianten in genen die betrokken zijn bij **synaptische** processen in **neuronen** waarin de **neurotransmitters** dopamine, acetylcholine en glutamaat een rol spelen.

Hoe veel **statistische power** we hebben om geassocieerde SNPs te vinden met GWAS, wordt onder andere beïnvloed door de nauwkeurigheid waarmee het bestudeerde fenotype is gemeten. Over het algemeen wordt in GWAS van ziekten klinische diagnose als **dichotoom fenotype** gebruikt (ziek of gezond). Maar een klinisch ziektebeeld (bijvoorbeeld stress-gerelateerde psychopathologie) is vaak heterogeen en een diagnose is geen duidelijke maat voor onderliggende biologische processen. In ***hoofdstuk 3*** zochten we naar SNPs en genen onderliggend aan stressregulatie en de ontwikkeling van gerelateerde psychopathologie in een militair cohort dat was uitgezonden naar Afghanistan. Daartoe implementeerden we meerdere **longitudinaal** gemeten, stress-gerelateerde, **kwantitatieve fenotypen** (metingen van stresshormonen en psychiatrische vragenlijsten vóór en na uitzending) in een GWAS. We ontdekten één locus wat we konden koppelen aan twee nabijgelegen genen (*INTS8* en

Psychiatrische en neurologische ziekten bestudeerd in dit proefschrift

Dit proefschrift bevat genoomwijde studies van verschillende complexe psychiatrische en neurologische aandoeningen, welke hier kort worden geïntroduceerd:

Schizofrenie (hoofdstuk 2 en hoofdstuk 6) is een relatief veelvoorkomende heterogene psychiatrische aandoening die gekenmerkt wordt door positieve symptomen (zoals hallucinaties), negatieve symptomen (zoals sociale terugtrekking) en cognitieve problemen (zoals vergeetachtigheid). Behandeling met antipsychotica is effectief in de bestrijding van positieve symptomen, maar de ontwikkeling van nieuwe medicatie wordt beperkt door gelimiteerde kennis over de oorzaak van de ziekte.

Stress-gerelateerde psychopathologie (hoofdstuk 3 en hoofdstuk 4) omvat de psychische problemen die ontstaan na het meemaken van extreme stress of een traumatische gebeurtenis. Vaak resulteert dit in een klinische diagnose met **posttraumatische stressstoornis (PTSS)** of **depressie**. Bepaalde beroepsgroepen, zoals militairen, hebben een verhoogd risico op deze aandoeningen. Een verstoring van de **hypothalamus-hypofyse-bijnier-as (HPA-as)** en daaraan gerelateerde processen hebben waarschijnlijk een belangrijke oorzaak in de ontwikkeling van stress-gerelateerde aandoeningen.

Amyotrofische laterale sclerose (ALS) (hoofdstuk 6 en hoofdstuk 7) is een progressieve en dodelijke neurodegeneratieve ziekte waarbij de **motorische neuronen** afsterven en patiënten verlamd raken. Er bestaat medicatie die het verloop van de ziekte kan afremmen, maar genezing is op dit moment niet mogelijk. Patiënten overlijden meestal drie tot vijf jaar na het ontstaan van symptomen.

Epilepsie (hoofdstuk 7) is een van de meest voorkomende neurologische aandoeningen die wordt gekenmerkt door herhaaldelijke insulten (epileptische aanvallen). De ziekte is heterogeen omdat de oorsprong van de insulten in verschillende hersengebieden kan liggen. Behandeling is mogelijk met anti-epileptica of een operatie, al is dit niet in alle gevallen effectief of mogelijk.

TP53INP1). Deze genen lieten een sterke associatie zien in onafhankelijke GWASs van andere stress-gerelateerde fenotypen zoals slapeloosheid, hoge bloeddruk en algeheel welbevinden. Verder onderzoek kan aantonen wat de precieze rol van dit locus of deze genen is in stressregulatie en daaraan gerelateerde psychopathologie.

In ***hoofdstuk 4*** onderzochten we of het risico op het ontwikkelen van posttraumatische stressstoornis (PTSS) en depressie na uitzending voorspeld kan worden aan de hand van genetische risicofactoren voor deze ziekten. Hiertoe berekenden we, op basis van bestaande grote GWASs voor PTSS en depressie, **polygenetische risicoscores (PRS)** en testten we de associatie tussen PRS en de ontwikkeling van symptomen van deze aandoeningen in militairen na uitzending. Omdat PRS het gecombineerde effect van duizenden SNPs vangen (die afzonderlijk erg kleine risico-verhogende effecten hebben) reflecteren ze het algehele genetisch risico op een aandoening. Ook onderzochten we de interactie van genetisch risico met jeugdtrauma en trauma gedurende uitzending, omdat dit belangrijke modulatoren zijn voor de ontwikkeling van PTSS en depressie. Met de huidige datasets konden we de ontwikkeling van deze psychiatrische aandoeningen na uitzending echter niet verklaren. Dit geeft aan dat de huidige PRS nog niet accuraat genoeg zijn om vóór uitzending aan de hand van genetisch risico te voorspellen wie een grotere kans heeft op het ontwikkelen van psychische klachten.

Omdat veel ziekten een complexe genetische achtergrond hebben waarin veel varianten een rol spelen is het niet uitzonderlijk dat bepaalde varianten **pleiotrope** effecten hebben, waarbij één genetische variant geassocieerd is met meerdere ziekten. Dit impliceert dat ziekten genetische risicofactoren en onderliggende biologie delen, en daarom opent het mogelijkheden voor gecombineerde analyses van ziekten. Daarmee kan de **genetische correlatie** tussen ziekten worden beschreven en kunnen SNPs worden ontdekt die een rol spelen in twee of meer aandoeningen. In ***hoofdstuk 5*** beschrijf ik veelgebruikte methoden om genetische correlatie en pleiotropie te onderzoeken met GWAS-datasets en beschouw ik de huidige en mogelijk toekomstige toepassing van dergelijke methoden in drie klinisch relevante domeinen in de psychiatrie: diagnose, prognose en behandeling.

De psychiatrie en neurologie worden vaak als twee losstaande klinische gebieden gezien, ondanks de aanwezigheid van symptomatische overeenkomsten tussen ziekten. Zo ontwikkelt ongeveer 34% van de ALS-patiënten gedurende het verloop van de ziekte cognitieve symptomen en komen in families van ALS-patiënten vaker **psychose** en gerelateerde aandoeningen zoals schizofrenie voor. Deze klinische observaties en de beschikbaarheid van recente en grote GWAS-datasets waren aanleiding om mogelijk gedeeld genetisch risico tussen ALS en schizofrenie te onderzoeken in ***hoofdstuk 6***. Hier toonden we een genetische correlatie van 14% aan en vonden we loci die mogelijk met zowel ALS als schizofrenie geassocieerd zijn. Daarnaast lieten we zien dat ALS niet significant genetisch correleert met andere psychiatrische ziekten. Dit is de eerste keer dat een genetische overlap tussen een neurodegeneratieve en psychiatrische ziekte is beschreven en geeft aanleiding om verder onderzoek te doen naar het onderliggende biologische mechanisme van deze overlap. Bovendien maakt dit onderzoek inzichtelijk waarom het van belang kan zijn om psychische symptomen in ALS meer onder de aandacht te brengen in de kliniek.

Gelijkende genetische overlaptechnieken passen we toe in ***hoofdstuk 7***, dat de genetische correlatie tussen ALS en epilepsie beschrijft. Naast cognitieve klachten ontwikkelen ALS-patiënten ook regelmatig krampen en **fasciculaties**. Deze worden mogelijk veroorzaakt door **hyperexcitatie** van **motorische neuronen**. Hyperexcitatie speelt een grote rol in epilepsie, en anti-epileptische medicatie vermindert hyperexcitatie in motorische neuronen van ALS-patiënten. Een gecombineerde genetische analyse van beide ziekten kon daarom het inzicht in een gedeeld genetisch mechanisme vergroten. We vonden echter geen genetische correlatie, wat erop wijst dat hyperexcitatie in beide aandoeningen vermoedelijk door onafhankelijke biologische mechanismen veroorzaakt wordt.

Uit de in dit proefschrift beschreven onderzoeken kunnen de volgende conclusies worden getrokken:

-Genetische varianten geassocieerd met schizofrenie accumuleren in synaptische processen in vooral dopaminerge, cholinerge en glutamaterge neuronen.

-In een GWAS van meerdere stress-gerelateerde fenotypen, gemeten voor en na militaire uitzending, vinden we een geassocieerd locus in de buurt van twee genen (*INTS8* en *TP53INP1*) waarvan we associatie repliceren in GWASs van andere stress-gerelateerde fenotypen.

-De ontwikkeling van posttraumatische stressstoornis en depressie na uitzending kan met polygenetische risicoscores gebaseerd op bestaande grote GWASs van deze aandoeningen niet voorspeld worden, ook niet wanneer de interactie van genetisch risico met jeugdtrauma of trauma gedurende uitzending wordt getest.

-Genetische overlap is frequent aanwezig tussen psychiatrische ziekten en gecombineerde analyses dragen bij aan vooruitgang in diagnose, prognose en behandeling.

-Amyotrofische laterale sclerose en schizofrenie zijn genetisch gecorreleerd. Dit is de eerste en tot dusver enige beschreven genetische correlatie tussen een neurodegeneratieve en psychiatrische ziekte.

-Op basis van de nu beschikbare data is er geen aanwijzing voor een genetische correlatie tussen amyotrofische laterale sclerose en epilepsie.

A

Verklarende woordenlijst

Desoxyribonucleïnezuur (DNA): Groot molecuul wat bestaat uit twee strengen die een karakteristieke dubbele helix vormen en is opgebouwd uit nucleotiden. Drager van de 'genetische code'.

Enkel-nucleotide polymorfisme: Een variatie in het DNA die één nucleotide (bouwsteen) beslaat (bijvoorbeeld: ...ATCGG**C**T... en ...ATCGG**A**T...) en in > 1% van de individuen in een populatie voorkomt.

Erfelijkheid: (Genetische) overdracht van waarneembare en niet waarneembare eigenschappen van een organisme naar een volgende generatie.

Fasciculatie: Snelle, onwillekeurige samentrekkingen van spiervezels.

Fenotype: Een waarneembare eigenschap (of het totaal aan waarneembare eigenschappen) van een organisme.
> Dichotoom-: Een fenotype met twee waarden. In GWAS van ziekten wordt vaak ziek (of aangedaan) tegenover gezond (of controle) als dichotoom fenotype gebruikt.
> Kwantitatief-: Een fenotype uitgedrukt als een (gemeten) waarde of een getal, zoals lengte. Ten opzichte van dichotome fenotypen zijn deze vaak specifieker en nauwkeuriger.

Gen: Een biologische eenheid van het DNA die codeert voor bijvoorbeeld een eiwit en daarmee een biologisch proces aanstuurt.

Genetische correlatie: De mate waarin twee fenotypen onderliggende erfelijke factoren delen.

Genoom: De complete genetische samenstelling van een organisme.

Genoomwijde associatiestudie (GWAS): Een hypothesevrije studie waarin de associatie van (meestal miljoenen) enkel-nucleotide polymorfismen, verspreid over het genoom, met een gemeten fenotype wordt onderzocht.

Hyperexcitatie: Ongecontroleerde hoge activiteit van een zenuwcel.

Hypothalamus-hypofyse-bijnier-as (HPA-as): Het belangrijkste hormonale systeem betrokken bij de reactie op stress, bestaande uit de hypothalamus, hypofyse en bijnieren (die het stresshormoon cortisol maken).

Locus: Een gebied of positie op een chromosoom of gen, in de context van GWAS vaak een met het bestudeerde fenotype geassocieerd gebied in het genoom.

Longitudinaal: Gemeten op meerdere tijdspunten.

Neuron: Zenuwcel.
> Motorisch-: Een zenuwcel die een spier aanstuurt.

Neurotransmitter: Een signaalstof die vrijkomt in de synaps en waarmee zenuwcellen met elkaar communiceren.

Nucleotiden: De bouwstenen van het DNA, bestaande uit een suikergroep, fosfaatgroep en een van de basen adenine (A), thymine (T), cytosine (C) of guanine (G) die de genetische code vormen.

Pleiotropie: Het verschijnsel waarbij een gen, of een genetische variant, een invloed heeft op meerdere fenotypen.

Polygenetische risicoscore: Een score die het algehele genetische risico op een ziekte reflecteert omdat deze is opgebouwd uit de som van alle ziekte-geassocieerde genetische varianten die iemand draagt.

Psychose: Toestand waarbij iemand de grip op de realiteit verliest, vaak gekenmerkt door hallucinaties en waanideeën.

Synaps: De plaats waar twee zenuwcellen contact maken en met elkaar communiceren.

Statistische power: Het vermogen van een studie om een nulhypothese te verwerpen. In de context van GWAS: het vermogen om significant geassocieerde genetische varianten te vinden.

Appendix III

Dankwoord

Ruim vier jaren onderzoek komen tot een einde. De genetische factoren zijn uitvoerig aan bod gekomen. Daarom richt ik me hier tot de 'omgevingsfactoren': een groot aantal mensen die een onmisbare bijdrage hebben geleverd aan het tot stand komen van dit proefschrift.

Copromotor, promotoren en beoordelingscommissie

Beste Dr. Luykx, beste Jurjen, als directe begeleider en copromotor wil ik je als eerste noemen. Ik heb je leren kennen als iemand met een haast onuitputtelijke positieve energie op alle vlakken van het wetenschappelijk onderzoek. Dat werkt absoluut aanstekelijk en gaf me regelmatig een motiverende boost na een van onze werkbesprekingen. Dat je bijna wekelijks de tijd nam om de voortgang te bespreken, maar me vooral ook de ruimte gaf mijn eigen planning te maken, waardeer ik erg. Daarnaast weet je mensen te verbinden, waardoor met samenwerkingen vanuit verschillende invalshoeken hele veelzijdige projecten ontstaan. Ook je interesse in ieders leven buiten de drukte van het onderzoek zorgt voor een prettige sfeer.

Beste Prof. Dr. Veldink, beste Jan, onder leiding van je enthousiasme voor het onderzoek en kennis over de genetica is jouw groep voor mij een mooie en uitdagende plek geweest om mijn promotieonderzoek te kunnen doen en me te ontwikkelen in een vakgebied wat voor mij nog grotendeels onbekend terrein was. Je hebt heldere ideeën over wat je belangrijk vindt in het onderzoek, bent op zoek naar vernieuwing en ziet graag dat dingen tot op de bodem uitgezocht worden. Daarnaast geef je iedereen de kans en de middelen om de (soms pittige) methodologie in de complexe genetica onder de knie te krijgen, met waar nodig de juiste begeleiding. Je drive om de genetica van ALS te ontrafelen is inspirerend, en ik wens je daar met jouw huidige en toekomstige team alle succes mee.

Beste Prof. Dr. Kahn, beste René, wat meer op de achtergrond maar altijd belangrijk in het bewaken van de voortgang, of met een kritische noot die mij weer aan het denken zette over de ingeslagen weg met een project. Hartelijk dank voor je begeleiding de afgelopen jaren.

Leden van de beoordelingscommissie, Prof. Dr. W.L. de Laat, Prof. Dr. D. Posthuma, Prof. Dr. W. Cahn, Prof. Dr. J.P.H Burbach en Dr. B.P.C. Koeleman, hartelijk dank voor jullie tijd en interesse om dit manuscript kritisch te beoordelen en deel te nemen aan de oppositie.

Collega's en samenwerkingen

Vier jaar geleden kwam ik de neurogenetica groep binnen, die op dat moment net gestart was. Een groep waarin zowel experimenten op het lab als complexe analyses achter de computer een grote rol spelen. Ik zou me vooral met dat laatste bezig gaan houden. Als een nog onervaren bio-informaticus volgden daarom enkele maanden van 'trial and error' op de command line, waarin ik soms uren aaneengesloten naar een niet-werkend script zat te staren om er uiteindelijk achter te komen dat er ergens één spatie miste. Maar zoals alle begin moeilijk is, is ook alles te leren; met als resultaat van het leerproces dit proefschrift, waarin deze groep een heel belangrijk aandeel heeft. Want ondanks de grote variatie aan projecten waar iedereen aan werkt, is er altijd ruimte voor overleg, hulp, het uitwisselen van data en scripts, of lange neurogenetica-meetings op de vrijdagochtend om een oplossing te vinden voor actuele analyse-issues. Daarbij komt dan nog de ongedwongen en goede sfeer waarin samengewerkt wordt, inclusief jaarlijks terugkerende activiteiten zoals de BBQ, en nuttige maar vooral ook hele gezellige congresbezoeken. Mede-onderzoekers Annelot, Bochao, Fabrizio, Gijs, Joke, Kevin, Mark, Paul, Perry, Raha, Ramona, Sara, Rick en Wouter, bedankt voor alles wat het werken in deze groep leuk en veelzijdig heeft gemaakt. Aoife, Bas, Brendan, Femke, Hermieneke, Kristel, Lindy en Maarten, veel dank voor jullie inspanningen waardoor ik altijd bij mijn data kon en zonder problemen het cluster vol kon laten lopen met analyses. En degenen die altijd klaar stonden als ik me verdwaald voelde in het lab: Chantall, Erwin, Jelena, Peter, Raymond en William, bedankt voor al jullie onmisbare hulp aan de bench en het meedenken over de experimenten.

Mark en Erwin, paranimfen, onze gedeelde interesses op het gebied van sport, muziek of biertjes hebben er ongetwijfeld aan bijgedragen dat we elkaar ook buiten het werk beter hebben leren kennen. Niet te vergeten resulteren de vaak door jullie georganiseerde sociale activiteiten in gezellige avonden met de

groep. Over het entertainment na mijn promotieplechtigheid hoef ik me dus vast geen zorgen meer te maken.

Raakvlakken met de psychiatrie, neurologie en translationele neurowetenschappen: het zorgde ervoor dat ik bij allerlei seminars, meetings, en kleine en grotere samenwerkingen betrokken raakte. Bedankt aan iedereen die me verder heeft weten te helpen met interesse voor – of nuttige input in – mijn projecten, of voor zomaar een leuke wetenschappelijke discussie.
Een aantal van de grotere samenwerkingen wil ik in het bijzonder benadrukken:

Beste Dr. Vinkers en Dr. Boks, beste Christiaan, beste Marco, jullie enthousiasme over de psychiatrische genetica en het onderzoek in het algemeen zorgden voor goede samenwerking binnen veel projecten in dit proefschrift. Bedankt voor de toegankelijkheid voor overleg waaruit regelmatig een realistische kijk op de resultaten, een vernieuwend idee, of een alternatieve invalshoek resulteerde.

Beste Prof. Dr. Vermetten en Dr. Geuze, beste Eric, beste Elbert, ik heb de projecten binnen de militaire psychiatrie als uniek, uitdagend en veelzijdig ervaren. De relevantie van het onderzoek werd voor mij extra duidelijk op het internationale veteranencongres, waar ik ons project mocht pitchen en de veelzijdigheid aan onderzoek in het vakgebied ontdekte. Er liggen dan ook vast en zeker nog veel mooie uitdagingen in de komende jaren. Bedankt dat jullie de genetische data aan mij hebben toevertrouwd, voor de interesse in de analyses en de begeleiding die ik heb mogen ontvangen.

Beste Prof. Dr. van den Berg, beste Leonard, ook jij werkt vol overgave aan de zoektocht naar de oorzaak en behandeling van ALS en motiveert de grote groep onderzoekers die hier dagelijks mee bezig is. Ik ben blij hiervan deel uitgemaakt te hebben.

Dear Dr. McLaughlin, dear Russell, although based in Ireland, you were one of the first people I met during my first day at work. Overwhelmed with the complexity of some of the analyses during my first weeks as a PhD student, the calm and above all enthusiastic way in which you clearly explained the methodology certainly made a difference. I really enjoyed the scientific discussions, or having a beer when we met in Utrecht or at a conference.

Dear Daniel, dear Vinicius, thanks for your help in getting me up to speed with bioinformatics when starting the first project of my PhD.

Beste Remmelt, het was prettig om met jou samen te werken binnen twee projecten: altijd goede communicatie en de uitgewisselde kennis over genetica aan de ene kant en de fenotypische analyses aan de andere kant waren volgens mij voor ons beiden een verrijking.

Dear Bobby, Remi and Mark. It was a pleasure to work together on the epilepsy-related analyses, which made a valuable addition to the development of my skills in genetics.

Studenten/stagiaires Gert-Jan, Marloes, Kendrick en Marie. Bedankt voor jullie inzet, waardevolle bijdrage aan projecten, en originele vragen en ideeën.

Dear Véronne, dear Nefeli, you were often fascinated by how I could possibly do research while only sitting behind my computer and not going into the lab to gather data. As 95% of my PhD indeed happened in the office behind a computer, it was important and great to have such nice officemates. Thanks for the regular distractions from my analyses to talk about the non-work-related things in life or to complain about experiments that weren't going as planned. After more than four years, we're almost there. Good luck with the very final stretches of your PhDs, and let's celebrate this year!

Finally, thanks to all the staff, technicians, secretaries and PhD students who are (or were) part of the department of Translational Neuroscience. It was absolutely great and unique to work in an environment where so many different disciplines of neuroscience research are represented and where all research groups are mixed up throughout the offices. In addition, the great social atmosphere allowed to have a lot of fun outside working hours: lots of sports (mud masters, Amsterdam city swim and a triathlon) some very good borrels (special thanks to the borrel committee members, it was great to organize these borrels with you for a couple of years!), a freezing cold PhD camping weekend, lab days, and many other celebrations. All in all, a working environment that has been really important in finishing this thesis and that I'm certainly going to miss.

Vrienden en familie

Familie Schijven, familie van de Ven, vrienden, bedankt voor jullie oprechte interesse in waarmee ik me de afgelopen jaren bezig heb gehouden.

Adriaan, Alexandra, Anne, Daan, Deborah, Gertjan, Harriëtte, Iolanda, Jan-Willem, Klaas, Laurens, Lena, Léonie, Marloes, Peter, Rosanne, Sebastian, Sjoerd, Tim, Ton, Vivian, Willem en Yvette. Voor een aantal lag de intro van de Radboud Universiteit in 2008 (groepje 6!!) en de studententijd die daarop volgde aan de basis van onze vriendschap. Studeren in Nijmegen was met jullie de perfecte mix van hard werken en feest vieren, een tijd om nooit te vergeten. Met een door de jaren heen verder gegroeide groep is het dan ook niet gek dat we elkaar nog vaak zien en enorm veel leuke dingen ondernemen, met de weekenden weg als hoogtepunt. Op naar nog veel mooie momenten!

Geert en Annemiek, bedankt voor de gezellige avonden in Nijmegen of Den Bosch tot nu toe. Hopelijk gaan we er na deze promotiedrukte nog meer aan toevoegen.

Cas, Kai, Pieter, Theo, Willem, Thijs, Amy, Anouk, Iris, Kelly, Marlotte en Merel. Het is fijn om regelmatig terug te keren naar de Tilburgse roots voor een feestje en een gezellige avond met jullie, wat de afgelopen jaren ook voor de nodige ontspanning tussendoor zorgde. Mannen, onze creatieve uitspattingen rondom carnaval, weekenden op festivals of stedentrips – die altijd gepaard gaan met veel lol en uitermate flauwe grappen – maken het een mooie groep. Tegenwoordig misschien wat minder vaak dan vroeger, maar laten we dat soort activiteiten er vooral in houden! Willem, extra bedankt voor je illustratiekunsten en hulp bij de lay-out van dit proefschrift.

Dimphy, als vriendin van de familie heb je samen met mij, mijn zus en ouders al heel wat mooie en minder mooie dingen meegemaakt en sta je altijd voor ons klaar. Jouw aandeel in het bereiken van deze mijlpaal is dan ook heel belangrijk geweest. Bedankt voor alles!

Mijn Griekse schoonfamilie, Katerina, Georgos και Nikos: ευχαριστώ για την καλοσύνη, ζεστασιά και αγάπη που μου έχετε δείξει από τη πρώτη στιγμή που με γνωρίσατε.

Evi, we zijn qua carrière onze ouders achternagegaan in de gezondheidszorg, maar waar ik me op het lab of achter de computer verschuil, sta jij daadwerkelijk aan het bed van de patiënt. Veel respect voor het werk wat je doet. Samen met Tom heel erg bedankt voor jullie steun en interesse tijdens de afgelopen jaren, en de vele gezellige en ontspannen momenten tussendoor. En natuurlijk op naar de volgende feestelijke gebeurtenis dit jaar.

Lieve pap en mam, met jullie als ouders was opgroeien absoluut geen lastige opgave. Voor jullie onvoorwaardelijke steun, aanmoediging en enthousiasme tijdens mijn school-, studie- en promotietijd ben ik jullie ontzettend dankbaar. Ik heb, voorzien van wat advies hier en daar, zelf mijn keuzes kunnen maken om van jullie vervolgens altijd het vertrouwen te ontvangen dat ik ook iedere volgende stap weer aankon. Dit proefschrift is daarvan het ultieme bewijs.

Elisavet, lieve Elli, je bent waarschijnlijk de enige die echt weet met hoe veel toewijding ik de afgelopen jaren aan dit boekje heb gewerkt. Gelukkig ben je ook bekend in het onderzoek en weet je daardoor dat het niet altijd zonder slag of stoot gaat. In al die drukte zijn jouw liefde en zorgzaamheid onmisbaar voor me. Ik geniet ontzettend van alle kleine en grote dingen die we samen ondernemen, van een avondje uit in Den Bosch tot een verre reis naar Cuba. Je geweldige inspanningen om helemaal in het Nederlandse leven te integreren (zonder je Griekse temperament te verliezen), zijn heel speciaal en iets waar je trots op kan zijn. Ik kan niet wachten om ook alle volgende stappen in mijn leven weer met jou tegemoet te gaan.

Appendix IV

Curriculum vitae and list of publications

Curriculum Vitae

Dick Schijven werd op 14 september 1989 geboren te Tilburg, waarna hij in de nabijgelegen dorpen Udenhout en Berkel-Enschot opgroeide. In 2006 behaalde hij zijn havodiploma aan het Durendaelcollege in Oisterwijk, en in 2008 rondde hij aan deze school het atheneum af. Aan de Radboud Universiteit Nijmegen startte hij vervolgens zijn bacheloropleiding biologie, waarin hij zich oriënteerde in de richting van de humane en medische biologie. Na het voltooien van zijn bacheloropleiding in 2011 vervolgde hij zijn studie in Nijmegen met de master medische biologie en daarbinnen de specialisatie neurobiologie. Gedurende twee grote stages deed hij ervaring op in het onderzoek. Zijn eerste masterstage liep hij aan de afdeling laboratoriumgeneeskunde in het RadboudUMC, waar hij onder leiding van Dr. Marcel Verbeek meewerkte aan de ontwikkeling van een assay voor de detectie in hersenvocht van eiwitten die een rol spelen in de ziekte van Alzheimer. De tweede stage volgde hij aan de afdeling gezondheid van vrouw en kind in het academische ziekenhuis van de Universiteit van Uppsala in Zweden. Onder leiding van Dr. Jocelien Olivier en Prof. Dr. Inger Sundström-Poromaa werkte hij mee aan onderzoek naar de effecten van depressie en antidepressivagebruik tijdens de zwangerschap op moeder en nakomelingen. Ook na het behalen van zijn masterdiploma in 2013 bleef hij nog een periode als onderzoeksassistent werkzaam in Uppsala. Vanwege zijn interesse in de moleculaire en genetische kant van het onderzoek startte hij eind 2014 in de neurogenetica groep in het Universitair Medisch Centrum Utrecht zijn promotieonderzoek onder leiding van Prof. Dr. Jan Veldink, Prof. Dr. René Kahn en Dr. Jurjen Luykx. Hier ontwikkelde hij vaardigheden in de bio-informatica en statistische analyse van grote genetische datasets. Zijn onderzoek richtte zich op complexe psychiatrische en neurologische ziekten en de genetische overlap daartussen. Zijn interesse in biologie, genetica en data-analyse zullen ook in het vervolg van zijn carrière een rol blijven spelen.

Dick Schijven was born on 14 September 1989 in Tilburg, the Netherlands. After finishing secondary school in 2008, he started studying biology at Radboud University in Nijmegen. In 2011 he obtained his bachelor's degree, and continued his studies in Nijmegen in the master program medical biology (with a focus on neurobiology). During two large internships he acquired experience in research. His first internship was in the department of laboratory medicine at Radboud University Medical Center, where under the supervision of Dr. Marcel Verbeek he worked on the development of an assay for the detection in cerebrospinal fluid of proteins involved in Alzheimer's disease. His second internship took place in the department of women's and children's health at Uppsala University academic hospital in Sweden. Under the supervision of Dr. Jocelien Olivier and Prof. Dr. Inger Sundström-Poromaa he participated in research on the effects of depression and the use of antidepressants during pregnancy on mother and child. After obtaining his master's degree in 2013, he worked as a research assistant in Uppsala for another six months. His interest in the molecular and genetic aspects of research led him to start his PhD studies in the neurogenetics lab of the University Medical Center Utrecht, the Netherlands, under the supervision of Prof. Dr. Jan Veldink, Prof. Dr. René Kahn en Dr. Jurjen Luykx. Here, he developed skills in bioinformatics and statistical analyses of large genetic datasets. His research projects focused on complex psychiatric and neurological diseases and the genetic overlap between those. His interest in biology, genetics and data-analysis will remain important factors in his future career.

List of publications

* indicates shared first authorship; ** indicates shared senior authorship

Included in this thesis

Dick Schijven, Jan H. Veldink, Jurjen J. Luykx. **Genetic cross-disorder analysis in psychiatry: from methodology to clinical utility.** *The British Journal of Psychiatry (in press).*

Dick Schijven, Daniel Kofink, Vinicius Tragante, Marloes Verkerke, Sara L. Pulit, René S. Kahn, Jan H. Veldink, Christiaan H. Vinkers, Marco P. Boks, Jurjen J. Luykx. **Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling.** *Schizophrenia Research (April 2018) PubMed ID: 29653892.*

Dick Schijven*, Russell L. McLaughlin*, Wouter van Rheenen, Kristel R. van Eijk, Margaret O'Brien, Project MinE GWAS Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, René S. Kahn, Roel A. Ophoff, An Goris, Daniel G. Bradley, Ammar Al-Chalabi, Leonard H. van den Berg, Jurjen J. Luykx**, Orla Hardiman** & Jan H. Veldink**. **Genetic correlation between amyotrophic lateral sclerosis and schizophrenia.** *Nature Communications (March 2017), PubMed ID: 28322246.*

Dick Schijven, Elbert Geuze, Christiaan H. Vinkers, Sara L. Pulit, Remmelt R. Schür, Marie Malgaz, Erwin Bekema, Jelena Medic, Kendrick E. van der Kust, Jan H. Veldink, Marco P. Boks, Eric Vermetten, Jurjen J. Luykx. **Multivariate genome-wide analysis of stress-related quantitative phenotypes.** *Manuscript sumbitted.*

Dick Schijven*, Remmelt R. Schür*, Marco P. Boks, Bart P.F. Rutten, Murray B. Stein, Jan H. Veldink, Marian Joëls, Elbert Geuze, Eric Vermetten, Jurjen J. Luykx, Christiaan H. Vinkers. **The effect of genetic vulnerability and military deployment on the development of posttraumatic stress disorder and depressive symptoms.** *Manuscript submitted.*

Dick Schijven, Remi Stevelink, Mark McCormack, Wouter van Rheenen, Jurjen J. Luykx, Bobby P.C. Koeleman**, Jan H. Veldink**, on behalf of the Project MinE ALS GWAS Consortium and the International League Against Epilepsy Consortium on Complex Epilepsies. **Common genetic risk is not shared between amyotrophic lateral sclerosis and epilepsy.** *Manuscript in preparation.*

Not included in this thesis

Johannes Hebebrand*, Triinu Peters*, **Dick Schijven**, Moritz Hebebrand, Corinna Grasemann, Thomas W. Winkler, Iris M. Heid, Jochen Antel, Manuel Föcker, Lisa Tegeler, Lena Brauner, Roger A.H. Adan, Jurjen J. Luykx, Christoph U. Correll, Inke R. König, Anke Hinney**, Lars Libuda**. **The role of genetic variation of human metabolism for BMI, mental traits and mental disorders.** *Molecular Metabolism (June 2018) PubMed ID: 29673576.*

Caroline M. Nievergelt, Adam X. Maihofer, Torsten Klengel, Elizabeth G. Atkinson, Chia-Yen Chen, Karmel W. Choi, Jonathan R.I. Coleman, Shareefa Dalvie, Laramie E. Duncan, Mark W. Logue, Allison Provost, Andrew Ratanatharathorn, Murray B. Stein, Katy Torres, Allison E. Aiello, Lynn M. Almli, Ananda B. Amstadter, Soren B. Andersen, Ole A. Andreassen, Paul A. Arbisi, Allison E. Ashley-Koch, S. Bryn Austin, Esmina Avdibegovic, Dragan Babic, Marie Baekvad-Hansen, Dewleen G. Baker, Jean C. Beckham, Laura J. Bierut, Jonathan I. Bisson, Marco P. Boks, Elizabeth A. Bolger, Anders D. Borglum, Bekh Bradley, Megan Brashear, Gerome Breen, Richard A. Bryant, Angela C. Bustamante, Jonas Bybjerg-Grauholm, Joseph R. Calabrese, Jose M. Caldas-de-Almeida, Anders M. Dale, Mark J. Daly, Nikolaos P. Daskalakis, Jurgen Deckert, Douglas L. Delahanty, Michelle F. Dennis, Seth G. Disner, Katharina Domschke, Alma Dzubur-Kulenovic, Christopher R. Erbes, Alexandra Evans, Lindsay A. Farrer, Norah C. Feeny, Janine D. Flory, David Forbes, Carol E. Franz, Sandro Galea, Melanie E. Garrett, Bizu Gelaye, Joel Gelernter, Elbert Geuze, Charles Gillespie, Aferdita G. Uka, Scott D. Gordon, Guia Guffanti, Rasha Hammamieh, Supriya Harnal, Michael A. Hauser, Andrew C. Heath, Sian M.J. Hemmings, David M. Hougaard, Miro Jakovljevic, Marti Jett, Eric O. Johnson, Ian Jones, Tanja Jovanovic, Xue-Jun Qin, Angela G. Junglen, Karen-Inge Karstoft, Milissa L. Kaufman, Ronald C. Kessler, Alaptagin Khan, Nathan A. Kimbrel, Anthony P. King, Nastassja Koen, Henry R. Kranzler, William S. Kremen, Bruce R. Lawford, Lauren A.M. Lebois, Catrin E. Lewis, Sarah D. Linnstaedt, Adriana Lori, Bozo Lugonja, Jurjen J. Luykx, Michael J. Lyons, Jessica Maples-Keller, Charles Marmar, Alicia R. Martin, Nicholas G. Martin, Douglas Maurer, Matig R. Mavissakalian, Alexander McFarlane, Regina E. McGlinchey, Katie A. McLaughlin, Samuel A. McLean, Sarah McLeay, Divya Mehta, William P Milberg, Mark W. Miller, Rajendra A. Morey, Charles P. Morris, Ole Mors, Preben B. Mortensen, Benjamin M. Neale, Elliot C. Nelson, Merete Nordentoft, Sonya B. Norman, Meaghan O'Donnell, Holly K. Orcutt, Matthew S. Panizzon, Edward S. Peters, Alan L. Peterson, Matthew Peverill, Robert H. Pietrzak, Melissa A. Polusny, John P. Rice, Stephan Ripke, Victoria B. Risbrough, Andrea L. Roberts, Alex O. Rothbaum, Barbara O. Rothbaum, Peter Roy-Byrne, Ken Ruggiero, Ariane Rung, Bart P.F. Rutten, Nancy L. Saccone, Sixto E. Sanchez, **Dick Schijven**, Soraya Seedat, Antonia V. Seligowski, Julia S. Seng, Christina M. Sheerin, Derrick Silove, Alicia K. Smith, Jordan W. Smoller, Nadia Solovieff, Scott R. Sponheim, Dan J. Stein, Jennifer A. Sumner, Martin H. Teicher, Wesley K. Thompson, Edward Trapido, Monica Uddin, Robert J. Ursano, Leigh L. van den Heuvel, Miranda van Hooff, Eric Vermetten, Christiaan H. Vinkers, Joanne Voisey, Yunpeng Wang, Zhewu Wang, Thomas Werge, Michelle A. Williams, Douglas A. Williamson, Sherry Winternitz, Christiane Wolf, Erika J. Wolf, Jonathan D. Wolff, Rachel Yehuda, Kieth A. Young, Ross M. Young, Hongyu Zhao, Lori Zoellner, Israel Liberzon, Kerry Ressler, Magali Haas, Karestan C. Koenen. **Largest genome-wide association study for PTSD identifies genetic risk loci in European and African ancestries and implicates novel biological pathways.** *Manuscript submitted.*

Raha Pazoki, Bochao D. Lin, Kristel R. van Eijk, **Dick Schijven**, Sinan Gülöksüz, GROUP investigators, Jurjen Luykx. **Phenome-wide and genome-wide analyses of quality of life in schizophrenia.** *Manuscript in preparation.*